DOCUMENT RESUME

ED 216 689 IR 010 204

AUTHOR	Lancaster, F. W.
TITLE	Guidelines for the Evaluation of Training Courses, Workshops and Seminars in Scientific and Technical Information and Documentation.
SPONS AGENCY	United Nations Educational, Scientific, and Cultural Organization, Paris (France).
REPORT NO	UNESCO-SC-75/WS/44
PUB DATE	Aug 75
NOTE	108p.; Best copy available.
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Behavioral Objectives; Check Lists; Educational Objectives; *Evaluation Criteria; *Evaluation Methods; Guidelines; *Information Retrieval; *Program Effectiveness; Questionnaires; Seminars; *Training; Workshops
IDENTIFIERS	*Scientific and Technical Information

ABSTRACT

This report discusses and presents guidelines for the evaluation of specific educational offerings--i.e., courses, workshops, and seminars--in the field of scientific documentation and information retrieval. The topics covered in particular sections of the report are levels of evaluation, the nature and role of evaluation in education, approaches to evaluation in education, techniques for measuring reaction, learning acquired, and educational program results. The evaluation of specific educational activities and the design of evaluation instruments are also discussed in detail. Accompanying the text are ten tables, two figures, a 46-item reference list, and seven appendices. Among the appendices are a selective bibliography on evaluation in education, examples of questionnaires used in formative and summative evaluation of courses or workshops, a report on the eleventh annual clinic on library applications of data processing, and an example of an educational evaluation checklist for use by an independent observer of a particular activity. (JL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED216689

UNITED NATIONS EDUCATIONAL,
SCIENTIFIC AND CULTURAL ORGANIZATION

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

GUIDELINES FOR THE EVALUATION OF TRAINING COURSES, WORKSHOPS AND SEMINARS IN SCIENTIFIC AND TECHNICAL INFORMATION AND DOCUMENTATION

Prepared under contract for Unesco
by F.W. Lancaster

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Susanne Adlung

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SC-75/WS/44

Paris, August 1975

Reproduced by permission of
Unesco

FR010204

CONTENTS

		Page
I.	Introduction	3
II.	Levels of Evaluation	4
III.	Evaluation in Education	7
IV.	Approaches to Evaluation in Education	8
V.	Evaluation of Reaction	15
VI.	Evaluation of Learning Acquired	27
VII.	Evaluation of Behavioural Change	32
VIII.	Evaluation of Programme Results	33
IX.	Evaluating Specific Educational Activities	38
X.	Design of Evaluation Instruments	44
XI.	Conclusion	57
XII.	References	60
	Index	63
	Appendices	

TABLES

Table		Page
1.	Levels of Evaluation	6
2.	Classification of "Approaches" to Evaluation in Education	12
3.	Major Categories of Data in the Evaluation of Education	13
4.	Types of Data to be Collected in the Formative Evaluation of a Course	20
5.	Some Important Questions to be Answered in the Summative Evaluation of a Course	21
6.	Behavioural Objectives for a Course in Information Storage and Retrieval	24
7.	Specimen Page from Proposed Curriculum for a Graduate Programme in Library Science (CONACYT)	28
8.	Instructions to Participants in a Course on "Position Classification"	36
9.	Sample Page from Student Self-Assessment Questionnaire Used in a Course by the U.S. Civil Service Commission	37
10.	Sample of Data from a Course Evaluation Questionnaire	52

FIGURES

Figure		Page
1.	Course Evaluation Questionnaire as Used at the University of Illinois. Verso: closed-ended questions	48
2.	Course Evaluation Questionnaire as Used at the University of Illinois. Recto: open-ended questions	50

I. INTRODUCTION

Evaluation procedures should be built into the educational process from the early planning stages of a particular programme. As indicated by Hampton (1), "evaluation must occur before, throughout and beyond the entire programme". Any programme for education and training in documentation needs, therefore, to be evaluated at various stages or levels, as follows:

(1) evaluating the need for various types of activity in order to be able to establish meaningful priorities;

(2) evaluating the goals and objectives of proposed activities to determine if they are, in fact, realistic;

(3) evaluating plans for the achievement of stated goals and objectives;

(4) evaluating educational activities while in operation;

(5) evaluating these activities in retrospect to determine if they did in fact achieve their stated objectives.

The purpose of this report is to develop guidelines relating to the evaluation of specific educational offerings - courses, workshops or seminars - in the field of scientific documentation. These guidelines relate to the fourth and fifth stages of evaluation identified above, namely the evaluation of courses while in operation and the evaluation of these courses in retrospect. The purposes of this evaluation are twofold: (a) to improve the quality and relevance of these programmes while they are in progress, and (b) to provide data that will allow the organization responsible for planning the programme to decide how successful a particular activity has been, how far it has contributed to the attainment of stated objectives, and what might be done to improve educational offerings of this type in the future.

For the purpose of this report the following definitions, taken from Good's Dictionary of Education,* are adopted:

> Course. Organized subject-matter in which instruction is offered within a given period of time, and for which credit toward graduation or certification is usually given.
>
> Seminar. An instructional technique common in but not limited to higher education in which a group of students engaged in research or advanced study meets under the general direction of one or more leaders for a discussion of problems of mutual interest.
>
> Workshop. An instructional method in which persons with common interests and problems meet with appropriate specialists to acquire necessary information and develop solutions through group study; usually residential and of several days duration.

In terms of evaluation, however, these distinctions are not worth maintaining because, although minor variations may exist, the problems of evaluation are mostly common to all three types of educational offering. Throughout this report, then, we will use the term "course" generically to refer to any type of educational offering in the field of scientific documentation.

---

* Good, C.V. Dictionary of Education. 3rd ed. New York, McGraw Hill, 1973.

II. LEVELS OF EVALUATION

In considering the evaluation of any type of activity it is important to recognize that various possible levels of evaluation exist. Three major levels may be identified, as follows:

1. evaluation to determine how well the activity or service is satisfying the needs of those for whom it was designed. We can refer to this level as evaluation of effectiveness;

2. evaluation to determine whether or not these needs are being satisfied as efficiently and economically as possible. This is cost-effectiveness evaluation;

3. evaluation to determine whether or not the cost of a particular service or activity is justified by the benefits accruing to the user community. This is cost-benefit evaluation.

While this categorization is a valid and useful one, it should be cautioned that, in practice, it is frequently difficult to make clear-cut distinctions of this type. In particular, it is often quite difficult to distinguish cost-effectiveness analysis from cost-benefit analysis and, in fact, these terms tend to be used very loosely in the literature. Costs, performance (effectiveness), and benefits of a particular activity are very closely interrelated and a complete evaluation of an activity would take all of these aspects into account (i.e., it would be a cost-performance-benefits study).

The evaluation of the effectiveness of an operation may itself be divided into two levels:

(a) evaluation of how well the operation is carried out;

(b) evaluation to determine how the operation is carried out and whether or not the operation can be improved.

The first of these levels will be referred to as macro-evaluation. Macro-evaluation is merely concerned with the overall assessment of performance (i.e., how well the objectives of the activity are being satisfied). Macro-evaluation can be purely subjective, conducted by a survey of user opinions, or it can be objective and, perhaps, quantitative (as in a case in which we derive a score for the activity on some reasonably well established scale). But macro-evaluation in itself tells us only how well an activity is performing. It does not tell us why the performance is at a particular level, what factors contribute to establishing the performance at this level, what failures occur in the operation, or how the operation might be improved. Macro-evaluation alone is a relatively sterile task since, presumably, the principal objective of any form or evaluation is to produce improvements in the activity being studied.

If we wish to improve an activity we must undertake a more detailed level of analysis to determine how it is carried out, how much failure occurs, why this failure occurs, and what might be done in the future to raise the overall level of performance. This analytical level of evaluation can be referred to as micro-evaluation. Micro-evaluation is essentially diagnostic, the objective being to improve the performance of the activity we are studying.

The various levels of evaluation mentioned above are recapitulated in Table 1. In general, the sequence is one of increasing complexity, with cost-benefit analysis being the most difficult to carry out. These levels of evaluation are as applicable

to educational activities as they are to any other. To take a simple hypothetical case, suppose a local medical society sponsors a short seminar to educate all the physicians in a community in the use of some new therapeutic or diagnostic procedure. The effectiveness of the seminar could be evaluated, at the macro-evaluation level, by asking the physicians attending how satisfied they were with the presentation and how effective it was in informing them of new developments in these techniques. Alternatively, the impact of the seminar could be evaluated more objectively (e.g., by using some kind of test instrument to determine how well informed the physician was in this area before and after he attended). Micro-evaluation would go beyond this, however, and would seek to discover, in detail, why 38% (say) of those attending were not satisfied with the seminar or why it had no significant impact in increasing the knowledge of a certain proportion of those participating. The objective of the micro-evaluation would be to improve this seminar or other seminars of this type in the future. A cost-effectiveness analysis of this activity would attempt to determine whether the objectives of the programme (i.e., informing the physician community of a medical innovation) could have been achieved more efficiently or more economically in some other way (e.g., by a video-tape or slide presentation which could be made available to each physician in his own office). The cost-effectiveness of an operation can be improved in one of two ways: (1) by raising the level of effectiveness but holding costs constant, or (2) by maintaining roughly the same level of effectiveness while reducing the cost of achieving this level. The video-tape or slide presentation, in this hypothetical case, would be more cost-effective than the seminar if it could be shown to be less expensive to implement but equally effective in transmitting the needed knowledge or, alternatively, if it could be shown to be more effective as a means of transmitting this knowledge while costing no more than the alternative form of presentation. Finally, a cost-benefit analysis would be concerned with programme justification. In this hypothetical case, do the benefits of this particular educational endeavour outweigh the cost of providing it? One way of approaching this would be to consider whether the medical society should be spending its money on this enterprise or whether it would in fact be more valuable (in terms of benefit to members or to the community at large) to spend the money in some other way. Cost-benefit studies are concerned with value per dollar invested. But the long-term value of many activities (and education is certainly not an exception) is exceedingly difficult to express in absolute terms and it is for this reason that cost-benefit studies are highly complex undertakings if approached in a systematic way. In actual fact, most cost-benefit "analysis" is highly subjective. A decision is made by an individual, or group of individuals, that a particular programme is worth carrying out. That is, they decide that the benefits of the programme outweigh the cost of providing it.

This report is concerned with the evaluation of education and training programmes in the field of scientific and technical documentation and information. The major emphasis of the report lies in the area of the evaluation of programme effectiveness, but some reference will occasionally be made to other levels of evaluation discussed above.

Table 1
Levels of evaluation

1. Evaluation of effectiveness
   a) Macroevaluation
   b) Microevaluation

2. Evaluation of cost-effectiveness

3. Cost-benefit analysis

III. EVALUATION IN EDUCATION

A very large body of literature exists on the evaluation of education. This literature is too extensive to be reviewed in detail in this report. Moreover, much of it is of only peripheral relevance to our specific interests: it deals with the general philosophy of educational evaluation, with evaluation of complete educational programmes, or with evaluation at levels with which this report is not directly concerned (e.g., evaluation of elementary education). A very useful and full review of the literature relating to the evaluation of educational and related programmes has been prepared by Steele (2). Another useful review of the literature is provided by Bergman and Siegel (3). The present report is very much more restricted in scope. It deals with the continuing education of adults, specifically professionals already involved in some way with the handling of scientific and technical information, and it is concerned only with the evaluation of relatively short training courses, workshops and seminars in a restricted subject field (the transfer of scientific, technical and related information), although the specific content of an educational programme should have comparatively little direct influence on the evaluation methodologies adopted. The emphasis of this report is very much a practical one. It attempts to identify specific procedures and techniques relevant to the evaluation of the type of educational effort with which information services are involved, and to present some guidelines for the selection and implementation of these procedures. Although a considerable amount of literature has been reviewed in the compilation of this report, it cites only items most directly relevant to the specific objectives noted above. A more complete bibliography on the evaluation of education is given in Appendix 1.

The case for evaluation in education has been very well stated by Hampton (1):

> "Evaluation is one of the most neglected aspects of programming in continuing education today. Yet it has the potential of being one of the most powerful administrative tools available to educators. Most programme planners do a respectable job of planning and carrying out their plans to a high degree of proficiency. At this point, however, the programme development process tends to break down, for too often planners fail to review their programme adequately before directing their attention toward planning another programme. Information derived from reliable appraisals of educational efforts is essential to the educator in order to rationalize decision-making with respect to improving future programming efforts. Formal and informal judgements made about the programme should logically evolve into a plan of operation, which ultimately may affect programme process, teaching methods and programme accomplishments".

Hampton insists, and rightly so, that evaluation should be regarded as an integral part of the educational process rather than just an adjunct to this process. Evaluation is an ongoing activity and it is an activity that should be built into an educational programme from the very beginning (i.e., from the initial planning stages).

Knox (4) identifies four questions that need to be answered in making decisions relating to evaluation: (1) who needs to know how effective the programme is, (2) what do they need to know, (3) who can best provide this information, and (4) how can it be obtained most effectively and efficiently? These matters will be addressed at various points throughout this report.

IV. APPROACHES TO EVALUATION IN EDUCATION

Perhaps not surprisingly, there are many ways in which approaches to the evaluation of education have been categorized in the literature. It would merely be confusing to try to summarize all of these various forms of categorization here. However, some of the more important are worth our consideration. Perhaps the most important distinction to make at the outset is that between summative evaluation and formative evaluation. Summative evaluation, as its name implies, is essentially retrospective. At the end of a course (or other educational activity) an evaluation is conducted to determine how successful this course has been. Summative evaluation has value in that, if properly conducted (at the micro-evaluation level as defined earlier), it will reveal failures or limitations of the course and suggest ways in which this or similar courses might be improved in the future. Summative evaluation can lead to improvements in future educational endeavours but, clearly, it cannot improve the course that has already been completed. Formative evaluation, on the other hand, is conducted while the course or workshop is in progress, the objective being to improve this particular educational experience before it is completed. To take a simple example, suppose we have a course that lasts for three weeks. We can evaluate the ongoing course at the end of the first week and again at the end of the second. On the basis of feedback from participants, we may be able to make changes to the course in order to make it a more meaningful educational experience in the second and third weeks. That is, we may change the teaching approach or we may change the direction or emphasis of the course content. In the case of a course or seminar of one week duration we can obtain daily reaction from participants. It is obvious that formative evaluation of this kind is more feasible with courses that are of relatively long duration. There is little possibility for the formative evaluation of a one-day seminar. Even if it were feasible to obtain participant reaction in the middle of the day, it might not be possible for the teaching staff to react sufficiently quickly to make the changes that might be needed to improve the seminar in the second period. The distinction between summative evaluation and formative evaluation is, however, an important one. It is a distinction that appears to have been made first by Scriven (5). Formative evaluation is not necessarily purely subjective. It is perfectly possible to have a formative test of the learning acquired by students. Such a test might, for example, determine whether or not the students can demonstrate the achievement of certain behavioural objectives established for a course. The distinction between formative and summative evaluation is not entirely clearcut, however, because the summative evaluation of one course or workshop may legitimately be regarded as part of the formative evaluation of a series of courses or workshops.

Whether it takes formative or summative form, the major purpose of evaluation in education is, according to Steele (6):

> "to determine the effects of teaching under given conditions on the knowledge, attitudes, and behaviours of those being taught to provide a basis for improving, justifying, or discontinuing the teaching activity".

Another important consideration is the viewpoint of the evaluation. For whom is the evaluation being conducted? Evaluation can be conducted from a number of different viewpoints, the most important being:

1. the viewpoint of the participants;
2. the viewpoint of the lecturers;
3. the viewpoint of those responsible for planning, administering or funding the programme.

The teaching staff itself should be most interested in formative evaluation, the object being to improve the quality of programme content and their own teaching methods. Steele (7) has stated the need for formative evaluation in this way:

> "Not only must a constant check be kept on how participants are reacting, but evidence of negative reaction must be fed back into the programme immediately so adjustments can be made".

Course participants will also be concerned with formative evaluation; they will want to provide feedback to the teaching staff in order, where necessary, to change the direction or emphasis of the course and thereby make it more responsive to their own requirements. Participants will also want to know how well they are progressing in a course (in terms of their own objectives or the objectives set by the teaching staff) and how they might be able to apply the learning gained. As Knox (4) points out, the participant may also want to discover what he does and does not know relative to the subject of the course - so that he may concentrate on what he does not know. Potential participants in a programme, on the other hand, are likely to be most interested in summative evaluations of previous courses organized by the same body and in the same subject area or one closely related. Potential participants in UNISIST courses, for example, are likely to be interested in student evaluations of previous UNISIST courses. Government agencies and other employers may also find such evaluations useful in deciding whether to send participants to a particular course and, if so, which people to send.

Planners and administrators of educational programmes, on the other hand, are likely to want evaluations that are more comprehensive and more far-reaching. Although they may want to see evaluation results from individual courses, they will also need to evaluate complete programmes and they are likely to be concerned with the total impact of the programme on all participants.

Another way of looking at the evaluation of education is in terms of <u>what</u> is being evaluated. Steele (8) identifies three categories of evaluation when the subject is viewed in this way:

(1) evaluation of the students (e.g., in terms of grades);

(2) evaluation of the teachers (for salary increases, promotion, tenure);

(3) evaluation of educational programmes.

The organizations to whom this report is largely addressed will mainly be interested in the third of these levels: the success or impact of a particular course or of an entire programme of courses. They will not be interested in evaluating the students as such, since most will not be examining bodies or agencies of accreditation, but they might well be interested in the evaluation of students as a means of assessing the success of a programme. For example, one way of measuring the success of a course is in terms of the level of knowledge of the students before and after they participate. Evaluation of the teaching staff will be important to the organizers, not for the reward purposes identified earlier, but as part of the overall programme evaluation and in the decision as to whether or not to use the same lecturers again on other assignments.

Evaluation may be formal or it may be informal. Informal evaluation of an educational offering occurs all the time. A lecturer will solicit feedback from participants at a workshop or seminar, over coffee, at lunch, or on other occasions in which he mixes with students. Even if he does not actively solicit evaluative data of this kind, he is likely to receive it gratuitously. Whether or not they are asked to do so, students (consciously or unconsciously) will conduct their own

evaluation of any course or workshop they attend. That is, they will make their own judgements about the quality and relevance of the content, the success of the approaches to teaching, and so on. By mixing with students an independent observer may be able to obtain a fairly clear idea of how the course is going. The teacher, also, is likely to sense how well a course is going and will be evaluating himself as the course progresses. In contrast to this informal or casual evaluation, formal evaluation implies a deliberate effort on the part of the organizers to get some assessment of the success of a course. A formal evaluation will involve some "standardized" approach to data gathering, usually by the use of a questionnaire, interviews with the participants, or some controlled comparison or test.

Another distinction worth making is that between internal evaluation and external evaluation. Internal evaluation is performed by those who conduct the educational programme and it is usually formative. External evaluation is performed by individuals somewhat removed from the programme. External evaluation is mostly summative.

There is still another classification that is worth mentioning: the distinction between subjective and objective evaluation. The subjective evaluation of an educational activity is usually much more easily accomplished than an objective evaluation. Subjective evaluation is based on opinions, the opinions of students, of teachers, or of an independent observer. Objective evaluation, on the other hand, attempts to move away from opinion, pure and simple, and to come up with an assessment that is more standardized and perhaps more quantifiable. An obvious example of an objective evaluation is one in which the success of a course is measured by testing the knowledge or abilities of the students before and after their participation. Some standardized test is applied before the course and again when the course is completed (the same test may be used before and after but, for reasons to be discussed later, it is preferable to use two different tests that are in some sense comparable), the purpose being to measure the change in the students as a result of their participation. Presumably, if the educational experience has been successful, the students will achieve significantly higher "scores" on the second test than they did on the first. A variation on this is the use of some standard test applied to two matched groups of students, each group having been exposed to a different method of presenting the same material, the object here being to compare the success of one approach with that of the other.

One of the clearest expositions of approaches to evaluation is that given by Hampton (1), who identifies four possible "steps" proposed earlier by Kirkpatrick (9). These steps are:

(1) evaluating reaction of participants;

(2) evaluating learning acquired;

(3) evaluating behavioural change;

(4) evaluating programme results.

Evaluation of reaction (of students, lecturers, observers) is easiest to accomplish. Such evaluation is completely subjective, although the data may be gathered in some systematic way and in a consistent form. These data may also be quantifiable in some sense (e.g., 80% of the participants were satisfied with the approach used in teaching). Evaluation of learning, on the other hand, can only be achieved through some objective procedure, usually some form of test. More difficult still is the evaluation of behavioural change in the participants. This goes beyond learning as such into the application of the learning acquired. It is possible for someone to "learn" a particular skill (in the sense that he can pass some test which may be

based on memorization) but still not be able to apply it in a practical situation. One approach to the evaluation of behavioural changes is by the measurement of on-the-job performance of an individual before his participation in some educational programme and again some time after his participation. Preferably this evaluation should be objective. Conceivably, however, it could be subjective, based on the opinions of his peers or his supervisors. Another possible method of measuring behavioural change is by the use of tests of problem solving or decision-making abilities. Programme evaluation differs from the types mentioned earlier in scale rather than in approach or form. The evaluation of an educational programme is of concern to those who plan and administer it. Programme evaluation implies the existence of a set of programme objectives. The evaluation is conducted to determine how well these objectives have been met. Clearly, a complete programme evaluation could involve studies of reaction, learning or behavioural changes, or all three of these, depending on what the objectives of the programme happen to be.

Knox (10) uses a somewhat similar classification of elements that may be evaluated: (1) general impact, (2) participant satisfaction, (3) content mastery, (4) personality, (e.g., improved self-image or greater social awareness), and (5) work-related (e.g., on-the-job performance).

Finally, as Hampton again points out, we can categorize evaluation activities by the stage at which these activities are performed, as follows:

(1) pre-course evaluation,

(2) ongoing evaluation,

(3) terminal evaluation, and

(4) follow-up evaluation.

These evaluation stages are obviously closely related to and identifiable with the various "approaches" mentioned earlier. Ongoing evaluation is likely to be formative while terminal evaluation is, by definition, summative. Follow-up evaluation is likely to be used in the measurement of behavioural changes, while the ongoing evaluation and the terminal evaluation are likely to be based on measurements of reaction and learning. Pre-course evaluation may be conducted for the purpose of evaluating learning or it may be used as one element in a more complete programme evaluation (i.e., some form of pre-course evaluation, in this case, may be conducted in order to establish meaningful objectives for the programme). Table 2 gives a summary of various approaches to the categorization of evaluation in education.

So far in this report we have largely restricted our discussion to judgemental aspects of evaluation; i.e., evaluation that allows us to make certain judgements about a programme and how far it has succeeded. Stake (11) (12) (13), however, has pointed out that descriptive data are also needed in the evaluation of education. The kinds of data to be collected on some educational programmes are viewed as a two-dimensional matrix, as illustrated in Table 3. Antecedents, transactions, and outcomes are various characteristics of an educational programme for which we may want to collect data. Antecedents are conditions existing before the educational programme, conditions that may affect the outcome, especially the characteristics of the students themselves - their experience, knowledge of the subject-matter, aptitudes, and so on. The antecedents are really the inputs to the educational activity, including instructors, materials and planning as well as the students. Transactions are the activities occurring during the programme, the encounters among students and teachers, students and students, students and materials, while outcomes are the results of the transactions, short term and long term, especially evidence of the impact of instruction upon the participants. These three characteristics (other

Table 2

Classification of "Approaches" to Evaluation in Education

By purpose	By viewpoint	By object of evaluation	By form	By objectivity	By characteristic measured	By stage
Formative	Student	Student performance	Informal	Subjective	Reaction	Pre-course
Summative	Teacher	Teacher performance	Formal	Objective	Learning	On-going
	Program planner	Program performance			Behavioral change	Terminal
					Program results	Follow-up

writers, Roid (14) for example, refer to these as inputs, processes and outputs) may be depicted by descriptive data or by judgemental data. According to Stake, intents are conditions that are planned for by the organizers or the instructors, including course content, teaching methods, and intended student behaviour (i.e., what teaching and what learning are intended). Observations are descriptions of what actually did occur, relating to antecedents and outcomes as well as transactions. Judgemental data are also of two types. We can judge various aspects of the programme according to some absolute standards that we adopt from elsewhere or devise for ourselves. Alternatively, we can use relative standards whereby we compare one programme with another or with a whole group of other programmes.

Stake points out that descriptive evaluation data can be used in two ways: (a) looking for contingencies among antecedents, transactions, and outcomes, and (b) looking for congruency between the intended and the observed. These uses of evaluation data improve our understanding of the effectiveness of the planning and implementation of a programme.

Provus (15) has described an evaluation model very similar to that of Stake. He recognizes four stages of evaluation corresponding to four stages of programme development: definition, installation, process and product. Evaluation moves through these four stages and through three major content categories: inputs, processes and outcomes (which correspond closely to Stake's antecedents, transactions and outcomes). The definition stage is concerned with programme intents. At the installation stage the evaluator is primarily concerned with observing transactions. At the process stage he responds to observed outcomes and tries to relate the outcomes to the transactions. Finally, at the product stage, he looks for congruencies between the intents and the outcomes.

Elsewhere, Stake (16) has made a distinction between "preordinate evaluation" and "responsive evaluation". Preordinate evaluation is based upon pre-specification in the form of statements of objectives, tests of one kind or another, and perhaps standards of various types. Responsive evaluation, on the other hand, is oriented less to programme objectives than to actual programme activities (the transactions of Table 3). As described by Stake, responsive evaluation is less formal than preordinate evaluation. The evaluator observes the programme in action and interacts with the various groups having an interest in the programme. Responsive evaluation is more descriptive than judgemental.

A categorization similar to that of Table 3 has been made by Knox (10). The key components in his model are: context (the setting and rationale of the programme), inputs (both people and materials), process and activity (which correspond to the "transactions" of Table 3), outcomes, judgements and application of findings.

This section of the report has described several approaches to the classification of evaluation in education. These various classifications, summarized in Table 2, are all closely interrelated (e.g., the purpose or viewpoint will influence the characteristic to be measured and purpose and characteristic to be measured will both determine the stage at which the evaluation is conducted). Table 2 is by no means a complete "classification of classifications". Alternative forms of categorization are possible and many of these alternatives have appeared in the literature. Nevertheless, this table seems to cover the most important ways of looking at evaluation, at least for the purpose of this particular study. The remainder of this report will mainly be concerned with methods of evaluation, relating these methods primarily to the characteristic to be measured.

Table 3

Major Categories of Data in the Evaluation of Education
as Identified by Stake

	Antecedents	Transactions	Outcomes
Descriptive data a) Intents b) Observations			
Judgmental data a) Absolute standards b) Relative standards			

(2) there is sometimes some doubt as to whether a respondent has answered truthfully or accurately and there may be no convenient or practical way to check the validity or accuracy of a response.

These objections are unlikely to be too serious in the situation under consideration here, namely the evaluation of educational offerings in the field of documentation. The number of people attending most of these programmes will be sufficiently small (indeed participation will usually be controlled to keep the groups small) that it will be possible to have someone in attendance while questionnaires are being completed within the group. This person (one of the lecturers perhaps) will be available to "interpret" the questionnaire to participants and to answer any questions they may have on how to complete it.

The question of veracity or validity of responses is unlikely to apply to a student evaluation of a course. This is a problem that is more likely to apply in a situation in which the respondent is, in some sense, evaluating himself. For example, there may be a tendency for a respondent to overestimate the number of journals he reads, the number of hours he spends in reading, or the number of his own publications. This is a matter of prestige and of the desire of the respondent to appear "in the best light". There is no reason to suppose, however, that the participant in a course would have any reason to be untruthful or otherwise inaccurate in the completion of a questionnaire - the nature of the questions asked virtually precludes the possibility of this problem occurring.

Interviews seem to have three major advantages over questionnaires as methods of data gathering:

(1) the presence of the interviewer tends to ensure that all questions are correctly interpreted by the respondent;

(2) it may be possible, by means of "probing" questions, for the interviewer to check on the accuracy of responses;

(3) the interviewer may be able to collect unsolicited observations from the person interviewed. Data unanticipated in the interview schedule may thus be collected.

We have already indicated that the first two of these advantages are unlikely to be very significant in the evaluation of programmes in this field, and the third benefit does not seem sufficiently important to warrant the use of interviews in place of questionnaires. Interviews are more expensive and time-consuming. They require scheduling of participants, which may not be at all easy to arrange, particularly at the end of a course when most of the attendees are likely to be anxious to leave. Moreover, interviews cannot be completed anonymously, unlike questionnaires, and, for reasons stated earlier, they would require use of an independent interviewer. Members of the teaching staff should not conduct such interviews: they are unlikely to get completely candid responses and they may, perhaps completely unwittingly, influence the responses by the way they pose the questions. For formal measurement of student reaction, then, the printed questionnaire is likely to be the preferred instrument for gathering data.

To allow meaningful evaluation, on the basis of reaction data, it is important that some statement of the objectives of the course be available to participants. One important facet of the evaluation will be the students' assessment of how well the course has met its stated objectives. If no stated objectives exist, or are available to students, the possibility exists that they will evaluate the course in relation to goals that the organizers never intended. While it will also be

V. EVALUATION OF REACTION

The most important aspect of this form of evaluation is presumably the reaction of the subjects of the educational activity; i.e., the "students" attending the course, seminar or workshop. Studies of participant reaction will tend to be subjective. That is, we use various methods to determine the opinions of the participants relating to the course in general and, possibly, certain specific features of it. At the most general level, an evaluation of reaction seeks to determine how "happy" the students are with the way a course is progressing or with the way the course was conducted. In fact, the type of data collected in this form of evaluation has been referred to as "happiness data" (see, for example, Knowles (17)) or as a "happiness index".

Evaluation of student reaction has definite value. As Hampton (18) points out:

> "It is important to know how people feel about the programmes they attend, for it is reasonable to expect that participants who enjoy a programme are more likely to obtain maximum benefit from it".

Knowles (17) has stated that:

> "On the whole, this kind of feedback is most useful in providing a general sense of trends in morale and satisfaction, but it frequently turns up specific and practical suggestions for improvement in the general programme or in specific activities; and it may reveal problem points that call for deeper evaluation".

"Reaction data" can be gathered for the purpose of formative evaluation or they can be gathered for summative evaluation. Both formal and informal procedures can be applied. At a more informal level, the teaching staff can ask the students for their unstructured, "off the cuff" impressions of the course. With a relatively small group (a group of the size likely to be typical of many UNISIST programmes, for example), this could possibly be achieved through an informal discussion with the entire group at the end of each day of meetings. At an even more informal level, it is quite likely that useful reaction data can be obtained through discussions, between students and teaching staff, over coffee, at lunch, and during organized social events. An independent outside evaluator can obtain reaction data in a similar informal way, by observing the progress of a course and by generally "mixing" with students. The outside observer may, in fact, be in a much better position than the teaching staff to gather informal reactions in this way. Participants who may be reluctant to criticize an instructor to his face may be more candid when asked for their opinions by some apparently impartial observer.

A more formal approach to the collection of reaction data will use some form of structured instrument for data gathering. Usually this will be a questionnaire completed by each student, probably anonymously, although interviews may be used in place of the questionnaire. If interviews are used, it is important that they be conducted in some consistent manner, following an interview guide of some kind. Questionnaires, although widely used and accepted as survey instruments in social science research, are criticized by some investigators, usually for two major reasons:

(1) questions may be misinterpreted by respondents and it is sometimes difficult to know whether or not the respondent has interpreted a particular question in the way the designer of the instrument intended;

important to learn of the personal objectives of students, and their judgement on how far these personal objectives have been met, a wide discrepancy between the objectives of the students and the objectives of the organizers would tend to indicate a failure in the advertising and promotion of the course or, possibly, the selection of students to participate.

But evaluation should not be solely on the basis of stated objectives. A course may have benefits to the participants that were not anticipated by the organizers. In some programmes, for example, the value of the social contact among students may be greater than the value of the formal presentations. It is possible that a course might receive a rather poor score in relation to its stated objectives but still be a valuable educational experience for some other reason. It is important, therefore, that the evaluation of any educational enterprise should be sufficiently open-ended that it will take account of benefits that are apparent to the sudents although not anticipated by those responsible for planning and organizing the endeavour. Parenthetically, it is worth mentioning that certain educational experiences may also have undesirable consequences that were unanticipated by the organizers or the teachers. For example, it is conceivable that a course on "literary appreciation" may cause certain students to hate reading, especially if the course requires the student to read a large body of literature in which he may have little or no interest. Harmful effects of this kind are more likely to occur in the basic education of children or young people. They are unlikely to occur in the continuing professional education of an adult population of the type this report is concerned with.

As stated earlier, studies of participant reaction may be used in both formative evaluation and summative evaluation. It may be worth while at this point to consider the kinds of questions that may be asked of students in each of these types of evaluation. More specific consideration will be given to the subject of questionnaire design later in this report.

Let us first consider the subject of formative evaluation and the kinds of information we need to gather, by questionnaire, at the end of each day or at the end of each week for the duration of the course. The following are probably the most important categories of information we need to collect from students for the purpose of formative evaluation:

1. The importance and relevance of the subject-matter covered up to this point in the course. If the course can be divided up into a series of discrete sections, we will probably want an assessment of the relevance of each section.

2. The quality of the teaching in terms of how the material has been presented. If a number of different lecturers are involved we may wish the students to rate the performance of each, particularly if the same teaching staff is to perform again later in the course or if we plan to use these lecturers in future courses.

If a number of different teaching approaches or teaching materials have been used in the course (e.g., films, transparencies, demonstrations, lectures, discussion groups), we will certainly want the students to indicate how successful each has been as a means of presenting the information.

3. The novelty of the information presented. Except in the case of a "refresher" course, a student will normally attend an educational offering in order to learn something new. We should therefore be interested in discovering how much the student is really learning (i.e., what proportion of the material he has been exposed to is new to him). Both relevance and novelty are important in the evaluation of teaching. The subject-matter of a course may be relevant to a student's interests but not new to him or it may be new to him but not directly relevant to his needs.

4. The "level" of the material presented. It will be important to know, as a course is progressing, whether the material presented is at a level appropriate for the particular audience addressed. It should not be too simple. The students must not feel that they are being "talked down to". Simplicity may also be related to novelty - the students may be learning nothing new because the lecturers have underestimated their prior knowledge of the subject-matter and are treating the material at too basic a level. Neither, of course, should the material presented be too complex for the audience (i.e., it should not be "over their heads"). This matter of level of treatment is extremely important in formative evaluation. We must know, early in the course, whether or not we are "reaching" our audience. There is no point in waiting until the end of the course only to discover that the majority of the participants have been completely bored, either because the material has been redundant or because they were unable to follow it.

The problem of "level" is likely to be a critical one in the presentation of workshops, seminars and courses within the UNISIST programme, for example, for the simple reason that the participants in any one of these activities are likely to vary widely in their backgrounds and experience. The fact that they come from different cultural backgrounds and different educational systems, together with the fact that their native tongues may not be the language in which the course is presented, makes the situation unusually difficult. The problem is compounded by the fact that the lecturers will usually not have met the students before the course begins and it may take a little time for them to find out exactly what the students already know about the subject-matter.

With the type of audience that can be expected for most programmes, in this field, it may be extremely difficult to conduct a course at a level that will be appropriate to all. It may be necessary to assume, at the outset, a minimum of knowledge on the part of all participants and to use the early part of the course to bring all participants up to about the same level, even if this means that some of them will find this early material redundant.

5. How well the student feels that he or she is progressing toward satisfying his or her own objectives in attending the course.

6. An indication of which features of the course, as conducted so far, have been most valuable or interesting and which have been least valuable or interesting.

7. If outside projects or readings have been used in the course, the students' assessment of the value and relevance of these experiences.

8. Any suggestions the students have as to how the course might be changed, in later days, to make it a more valuable educational experience.

9. The overall rating of the course, on some scale, by all participants.

10. Any other observations or comments the students would like to make.

A formative evaluation does not necessarily have to be as highly structured as we have suggested. There may be something to be said for a more free-form approach. Woodward and Yeager (19), for example, used daily logs in which students record general impressions of each day in an anecdotal, free-response form. In this they were guided only by very general headings such as "problems encountered". Moreover, a formative evaluation does not necessarily have to be done on a daily or weekly basis. Instead a formative questionnaire may be developed for each section of the course and administered when each section is completed.

In all of this data gathering it will be important to discover why various responses are made. If students express dissatisfaction we must learn the precise nature of this dissatisfaction if we are to make changes designed to improve the course. Thus, we must discover exactly why a student feels he is not making satisfactory progress in the course, which parts of it he has found too complex, which parts repetitious, and so on. This implies that the questionnaire should be designed in such a way that, whenever the student indicates some form of dissatisfaction, he is asked to describe the precise nature of this dissatisfaction.

In the above discussion we have noted the kinds of data we will need to collect in order to conduct a formative evaluation of some educational offering. These data are summarized in Table 4. The tabulation is undoubtedly incomplete. In particular, each course or workshop may have some special features that require emphasis in the evaluation questionnaire. Nevertheless these data seem to be those of greatest overall importance and they should be generally applicable in the formative evaluation of most of the activities with which organizers of courses in this field will be involved. Some examples of questionnaires used in formative evaluation are given in Appendix 2. A more detailed discussion on the design of questionnaires will appear later in the report.

In the summative evaluation of the course we may want to ask some of the same questions that were used in its formative evaluation, this time addressed to the course as a whole rather than just a part of it. We are also likely to include some types of questions that we did not use in formative evaluation. Some of the major questions to be answered in a summative evaluation are presented in Table 5. Most of these are self-explanatory. At this point it will be necessary to ask students to view the course in its entirety as an educational experience, including various factors that may have contributed to an atmosphere conducive to learning. These factors will include physical arrangements (the lecture and meeting facilities and hotel or other accommodations) and organized social events. In this summative evaluation we are concerned with looking back on the course, to determine how successful it was in toto and to identify any failures or problem areas that may have occurred. On the basis of this experience we may be able to make valuable changes to the course in the future and we can draw on this experience in the planning and execution of future courses in other subject areas.

An important element in the summative evaluation will be the determination of how well the objectives of the students have been satisfied. It will be important to determine, for each student, what his or her objectives were in attending the course, and how far these objectives were actually met. In this connexion, it might well be desirable to use some very brief "pre-course questionnaire" (perhaps incorporated into the actual application form) to determine student objectives before the course begins. At the end of the course, we can present to each participant the objectives he specified before the course began. He can then be asked (a) to modify this statement of objectives if he feels that, in retrospect, it was too narrow, too broad, or otherwise inappropriate, and (b) to judge the success of the course in meeting his original or revised objectives. It will also be desirable to include a statement of the objectives of the course as prepared by the organizers of the course and to ask students how far they feel that these general objectives have been met.

As in the formative evaluation, it will be important in summative evaluation to discover why, especially in situations in which the student has indicated dissatisfaction with some aspect of the programme. Some examples of questionnaires used in the summative evaluation of courses or workshops are given in Appendix 3.

Table 4

Types of Data to be Collected in the
Formative Evaluation of a Course

1. Importance and relevance of subject matter.
2. Quality of the presentation
    a) by individual instructor
    b) by form of presentation used.
3. Novelty of the information transmitted.
4. Intellectual "level" of the material presented.
5. Students' assessment of their own progress.
6. Most and least valuable features of the course so far.
7. Value of outside projects and/or readings.
8. Student suggestions for improvement of the course.
9. Overall student rating of the course.
10. Any other observations the student wishes to make.

Table 5

Some Important Questions to be Answered in the Summative Evaluation of a Course

1. Have students' objectives been met? Have overall course objectives been met?
2. How valuable has the course been to participants in terms of their own professional goals? What were the main benefits gained?
3. How do students rate the quality of the instruction
    a) by individual instructor
    b) by type of learning experience (lecture, discussion, practical sessions, and so on)?
4. Size of group participating--too large, too small, about right?
5. Was the level of treatment of the subject matter at the "correct" level for the majority of the audience?
6. Was coverage of subject matter comprehensive or were there notable gaps in coverage?
7. Were the lectures and/or other educational experiences well integrated into a meaningful and complete program, or were they disjointed and overlapping?
8. Was most of the material presented "new" to most of the participants, or was it repetitive and redundant?
9. Which features of the course were most valuable and which least valuable to those attending? Another way of putting this is in the form of features that participants would like to add to or delete from the course.
10. Were the physical and related arrangements (lecture halls, accommodations, meals and/or coffee breaks, social events) satisfactory?
11. Would participants recommend this course to their colleagues and would they attend another, similar course if it were offered?
12. Any other ways the course could be improved?

In a course of fairly long duration, in which a somewhat wide range of related topics is discussed, it will be necessary for the lecturers to develop a set of behavioural objectives for the course, with one or more behavioural objectives associated with each segment of the course. If, at the conclusion of the course, each student is asked to indicate how far he feels these behavioural objectives have been met, we can legitimately regard this as a part of our evaluation of reaction. If, on the other hand, we develop some "demonstration of achievement" for each of the behavioural objectives, and use some objective testing procedures to determine how far the behavioural objectives have been met, we are really measuring the learning of the students, and this form of evaluation then falls legitimately in the category "evaluation of learning acquired". This will be discussed in the next section of the report. In Table 6 are presented, by way of an example, some specific behavioural objectives, together with demonstrations of achievement, that were developed for the Unesco International Training Course on Information Retrieval and Information Retrieval Systems, held in Katowice, Poland, in August 1974.

So far we have discussed only the reaction of students. While student reaction is obviously of great importance, it is not the only reaction data of interest. Wherever it can be arranged, it may be highly desirable to have some independent observer present at some or all of the presentations. Such an observer can serve two functions: (1) to mix with students and obtain feedback and reaction that may be more "candid" and impromptu than the reaction data obtained by the teaching staff through more formal procedures, and (2) to observe the teachers in action and thus to arrive at his own assessment of the quality of the course, both in terms of its content and in terms of the methods of presentation. This second function may be regarded as a form of "peer review". It requires, therefore, the use of an individual who is knowledgeable in the subject-matter covered in the training programme and at approximately the same level of seniority and professional experience as the members of the teaching staff themselves. It may be desirable to provide this outside observer with some type of checklist on which he can record, in standardized form, his observations on course content, teaching methods, student interest, and whatever else we may be concerned about. An example of such a checklist is given in Appendix 7.

Another "reaction study" will involve the teaching staff. It will be valuable to have the lecturers evaluate the course from their point of view, and to prepare a summary of this evaluation once the course is concluded. A good lecturer will constantly be evaluating himself. He is likely to recognize the fact that certain material was not presented as clearly or as completely as he would have liked, that a different sequence of presentation would have improved the situation, that certain types of information could be better presented in an alternative way (e.g., by student readings and group discussion instead of straight lecturing), and that certain topics might be omitted entirely in the future, because they were redundant, of only marginal relevance or interest, or because they were clearly boring to the audience. The lecturers' evaluation should also include their observation on the students - their quality, degree of interest in the subject-matter, the intelligence of the questions asked, their diligence (e.g., in completing background reading and other assignments), and their general suitability for participating in a course of this type. It is possible that a course may fail to meet part of its objectives because some of the students did not have the background necessary to benefit fully from the experience. This could be a significant problem in the case of a programme that may be attended by participants from many different countries, with widely divergent backgrounds, and varying levels of competence in the language in which the course is offered. In any case, the lecturers' evaluation of the course, and of the student body, is likely to be an important element in a complete evaluation programme, and such an evaluation may have great significance in the advertising of future courses and in establishing appropriate criteria for the selection of candidates.

The teaching staff should also evaluate the "environment" of the course: the adequacy of the facilities, the helpfulness of the local organizers, and other factors that may add to or detract from the educational aspects.

Knowles (20) has mentioned that instructors have great shortcomings as observers because they "are personally involved in the outcome of the evaluation, so that it may be difficult for them to be objective. They may tend to overlook instances in which desired changes are not being produced and to emphasize minor successes". On the other hand, it could be argued that the reverse situation might also be true. Some instructors may be too sensitive, too self-critical. They may find faults where none really exist. Whatever the limitations of the instructor as an evaluator, however, it is clear that he has an important rôle to play. His input will be analysed and interpreted along with the input from students, independent observers and other individuals who are involved in some way with the educational experience. It is important that a complete evaluation should be based on input from a number of individuals representing different levels of involvement and points of view.

Table 6

BEHAVIORAL OBJECTIVES FOR A COURSE IN INFORMATION STORAGE AND RETRIEVAL

Objectives	Demonstration of Achievement
1. Understand the input and output operations involved in modern computer-based IR systems.	1. Be able to describe the functions of such a system in terms one of your own students would understand.
2. Recognize the advantages of on-line and off-line IR systems.	2. Be able to enumerate these advantages and discuss them clearly and concisely.
3. Understand how search strategies are constructed and understand the Boolean logic involved in such strategies.	3. Be able to conceptually analyse a request for information, represent the relationships graphically by Venn diagrams, reduce the conceptual analysis to a set of index terms and logical operators, and reduce eventually to a simple equation.
4. Understand the distinctions between pre-coordinate and post-coordinate retrieval systems and the advantages and disadvantages of both types.	4. Be able to categorize any system as pre-coordinate or post-coordinate. Be able to write a paper clearly distinguishing the two types, with examples, and comparing the pros and cons of each.
5. Understand the manual and semi-mechanized post-coordinate systems that were the forerunners of modern machine-based systems.	5. Be able to set up simple card files to demonstrate the peek-a-boo, uniterm and edge-notched card principles.
6. Recognize the principle types of tools or services by which a scientist may keep current in the literature of his subject field.	6. Be able to write a paper describing the current awareness services that you would provide were you appointed to the post of technical librarian or information officer in an industrial organization.

Table 6 (continued)

7. Recognize the distinction between social, archival and dissemination functions of the scientific journal.	7. Be able to enumerate and describe the limitations of the scientific journal as a dissemination mechanism. Be able to describe some "conceptual alternatives" to conventional journal publication.
8. Be aware of the major information centers existing in the United States and elsewhere and the principal types of service that these centers provide.	8. Be able to describe the principal operations and services of at least: NLM, NTIS, DDC, NASA, ERIC, SIE, the National Referral Center for Science and Technology, the "information analysis center", and the "scientific information dissemination center".
9. Recognize the principal requirements of a user of an IR system, and the criteria by which the user will judge the performance of the system.	9. Be able to enumerate these criteria and discuss the interrelationships and tradeoffs between them.
10. Understand the major factors governing the performance of any IR system.	10. Be able to list these factors in a table, or reduce them to the form of a diagram, and be able to discuss them. For a particular system of given characteristics, be able to identify the factors most likely to be influencing its performance.
11. Understand why, and in broad terms, how an information retrieval system is evaluated.	11. Be able to read and understand Evaluation of the MEDLARS Demand Search Service.
12. Understand, in general, the construction of controlled vocabularies for IR systems, with special reference to thesauri, and the influence of the vocabulary on the performance of such a system.	12. Be able to demonstrate the structure (BT, NT, RT relations) of a thesaurus. Be able to list and describe the major "index language devices". Be able to write a paper on the role of an index language in the performance of a retrieval system.

Table 6 (continued)

13. Understand, in general, how a computer may be used to: (a) search full document text; (b) index a document by extraction and assignment; (c) prepare an abstract; (d) form classes of terms or documents on the basis of statistical associations.	13. Be able to: (a) describe the major techniques used in "natural language searching"; (b) distinguish between extraction indexing and assignment indexing; (c) distinguish the relative frequency approach from the absolute frequency approach to automatic indexing; (d) define the concepts of "clumps" and "associative retrieval"; (e) summarize the major features of SMART.
14. Understand the importance of machine-readable data bases, the range of available data bases, and the effect of the availability of these data bases on the provision of information services.	14. Be able to prepare a plan for the implementation of an information service, using existing data bases, in your own country.
15. Recognize the major terminology and concepts in the field of information storage and retrieval.	15. Be able to read critically and understand major portions of the Annual Review of Information Science, the Journal of the American Society for Information Science, the Proceedings of ASIS, and the Journal of Documentation.

VI. EVALUATION OF LEARNING ACQUIRED

The evaluation of the learning acquired by students is more difficult than evaluation on the basis of student reaction only; it requires the use of more objective procedures. One possible approach, mentioned in the preceding section, is the use of behavioural objectives, with accompanying "demonstrations of achievement", as the basis for measuring the learning imparted. A sample set of behavioural objectives is shown in Table 6. As can be seen from this table, several of the "demonstrations of achievement" can be tested objectively. For example, to show understanding of the hierarchical and cross-reference structure of a thesaurus, the student can be asked to construct a sample of four or five thesaurus pages in a subject area of his choice. A complete curriculum may be constructed in the form of a collection of behavioural objectives. Table 7, for example, shows a portion of a curriculum in library science proposed for implementation in Mexico by the Consejo Nacional de Ciencia y Tecnologia (CONACYT). In developing sets of behavioural objectives of this kind, as pointed out by Hunter and Schooley (21) it is wise to order the objectives by internal priorities. For example, if achievement of objective B is first dependent on the achievement of objective A, then A possesses a higher internal priority than B.

Another, somewhat similar procedure would involve the use of a brief test administered to students before they participate in the course. Such a test could be made part of a pre-course questionnaire, which is also used to determine the students' expectations concerning the course. The test would be completed anonymously (to avoid possible embarrassment on the part of students, who might feel uncomfortable if they were unable to answer many of the questions posed). This type of test has value for two purposes:

(1) the teaching staff will gain a better idea of what the students, as a group, already know and what they do not know;

(2) the same set of questions can be used, when the course is completed, to test the students again and thus to measure the change in their knowledge, at least as far as this set of questions is concerned.

This technique of pre- and post-testing of student knowledge is likely to be most easy to apply where the questions can be made reasonably factual.

In actual fact, there is some possibility of bias associated with the technique as described above. If the lecturing staff prepare the set of questions there is a danger that, consciously or unconsciously, they will give particular emphasis to these questions in their actual presentations. If the same questions are then used at the end of the course the "evaluation of learning" may be somewhat biased in favour of the course since it is possible that other questions, although equally important, might not be answered nearly as well by the group of students. One way of avoiding this possible bias is to have the questions composed by an outsider experienced in this subject field (perhaps the person who was mentioned as an independent observer in our discussion of the measurement of student reaction), his questions being based on the course outline and its behavioural objectives. This set of test questions would be answered by students, to the best of their abilities, at the very beginning of the course, and would again be used at the end of the course. They would not, however, be seen by members of the teaching staff until after the course is completed.

This technique, although an improvement on the one first mentioned, also has some slight possibility of bias. It is possible that the students themselves, remembering the questions asked in the pre-course questionnaire, will themselves concentrate on these topics in their study, possibly to the exclusion of other topics

Table 7

Specimen page from Proposed Curriculum for a Graduate Program in Library Science. Reproduced by permission of the Consejo Nacional de Ciencia y Tecnologia (CONACYT), Mexico.

AREA: Administration and Research

COURSE: Mexican Academic Libraries

QUARTER: Third

Description: Discuss the role of the academic library, with emphasis on problems of organization, administration, service, and the relationship of the library to the overall educational program. Emphasizes the application of administrative principles to the practices and organizational patterns in junior college, college and university libraries.

Unit I. The Library -- Internal Organization 30%

Through reading assignments, class presentations and discussion, to compare and contrast:

1. Types of administrative organization and how they lend themselves to academic libraries. (functional, subject, divisional, etc.)
2. Centralization vs. decentralization of libraries (branch libraries).
3. Methods of control used to express achievements of the library; application of statistics to library operations.
4. Suggested standards for collection, space, staff, budget, evaluation of the total academic library, etc.
5. What problems are evident in the organization of academic libraries (closed vs. open stacks, control system, classification schemes, etc.)

Unit II. The Services -- Personnel 30%

Through the assignment of relevant resources recognize:

1. How personnel are recruited to libraries.
2. In-service training for all levels of library personnel.
3. What classification of positions and salaries is possible.
4. The interpersonal relationships that present particular problems in an academic setting.

Unit III. The Service -- Administrative 20%

1. What the sources of income are for academic libraries (institutional, grants, gifts, and exchanges, etc.)

of equal or greater importance. This could be true even when the student is given no indication that he will be tested again at the end of the course, a condition which is essential to the conduct of an evaluation of this type.

The possibility of bias in the evaluation results could be reduced by a slightly more complex experimental design. This would involve a "cross-over" test in which the set of questions, as well as the group of students, is divided up into two sets. Question set A is administered to student group A before the course and to student group B after the course, and vice versa for question set B, as shown below:

Student group	Pre-course	Post-course
A	A questions	B questions
B	B questions	A questions

It could be argued, however, that such a test might bias the results against the course because A students might again concentrate, in their study, on A questions and thus do unnecessarily poorly on the B questions at the end, and vice versa for the B students. It is also likely that students from both groups might discuss the test questions with each other, although this danger could probably be minimized by asking the students to treat the test questions as confidential.

It can be seen, from the above discussion, that it is somewhat difficult to arrive at a test design, for this situation, that has no possibility of bias in one direction or another. A better design would probably be one involving a control group of people who have not taken the course but who are otherwise well "matched" with the student group in terms of other characteristics, especially in their educational background and level of experience. The use of a control group in this way would eliminate the need for the use of pre-course questions. The two groups would simply answer the same set of questions when the course is completed, and the two sets of results can then be directly compared, the assumption being that the student group should get significantly better scores than the control group. This type of measurement of learning, although possible within a university environment, would be extremely difficult to implement with most courses, where potential groups of participants are much less "captive", and we will not give it further consideration for this reason.

Although we have taken pains to point out the possibility of various biases in approaches to the measurement of learning, these biases may be less serious in practice than the discussion implies. It is quite likely that the crossover test, as described above, would give a reasonably reliable indication of how much learning has actually occurred during a particular course, especially if the questions are composed by an independent evaluator.

The use of an independent evaluator for this purpose raises problems of its own. It is obviously extremely important that this evaluator should have a very clear understanding of what the course is intended to accomplish, otherwise his questions may test for information that the lecturers themselves consider outside the scope of the course. Alternatively, his questions may test in areas that the teaching staff considers of peripheral importance, to the exclusion of more important areas, or he may use terminology that differs from the terminology adopted by the lecturers, thus causing great confusion among the students.

Let us try to summarize the situation relating to the evaluation of learning acquired:

1. Two possible approaches can be used:

   (a) the development of specific behavioural objectives for various segments of the course and the use of "demonstrations of achievement" while the course is in progress, and

   (b) the use of a before and after test, again based on the behavioural objectives of the course.

2. If a before and after test is used, the students should not be told in advance that they will be tested at the end of the course.

3. It is highly desirable that a different set of questions be used for the "before" and "after" portions of the test.

4. This can be achieved in one of two ways:

   (a) the use of two sets of questions, of approximately equal complexity, one used before and one used after the course, all students answering all questions, or

   (b) a crossover test of the type mentioned earlier.

5. The lecturers can develop their own set of questions provided they are careful to ensure that the entire set of questions ranges over the complete subject-matter to be covered in the course.

6. Alternatively, the question set can be devised by an independent investigator and the test applied by him.

7. If an independent evaluator is used he must work closely with the lecturers before the course begins and there must be clear understanding between them on terminology, on the scope of the course, and on its behavioural objectives.

Mehlinger and Patrick (22) have pointed out that an evaluation instrument that tries to measure learning must satisfy three basic requirements in order to be valid. Test items must be specifically related to the stated objectives of the course. Experts must be able to agree on what is the "correct" or "best" answer to a question. And most students who have not taken the course should not be able to respond correctly to the test items (this implies the use of a control group).

Veri et al (23) refer to requirements for the evaluation of learning in the following terms:

> "An ideal procedure is to measure the cognitive, affective and psychomotor skills of the learners before they enter the learning process, teach to behaviourally defined objectives, and then measure what has been learned with somewhat detailed objective tests. The tests must be designed meticulously using accepted test construction methods and procedures. Follow-up testing at some later time to measure retention of learning completes the ideal model".

Clearly, it is easier to test learning acquired in a course if its main purpose is to impart factual information or if it is designed to teach a particular skill (e.g., the production of a thesaurus). In particular, a test of learning will be easiest to apply and to grade if it can be reduced to a set of questions of a multiple choice or a true/false type. This is not always possible. In fact, a course may be designed not to impart factual information but to deal with problem

solving or decision-making situations. It is more difficult to test the impact of a course on the problem solving or decision-making skills of participants but it is not impossible. One possible technique has been described by McGuire and Babbott (24). The test was devised to assess the abilities of medical students in taking care of a sick person. Each student is given a medical case history on the patient along with several possible diagnoses or courses of action. The student must choose among the alternatives presented and each choice branches to other choices. Eventually, the patient will either recover, be referred elsewhere, or will die. A panel of medical experts is used to rate each choice at each decision point on a five-point scale, ranging from "clearly indicated" to "clearly contra-indicated". An efficiency score, a proficiency score and a composite score are calculated for each student. The efficiency score represents the proportion of the student's responses that are beneficial to the patient and the proficiency score represents the percentage agreement between the student's choices and those of the panel of experts. The composite score combines the efficiency and proficiency scores. Although this test was designed for the health care situation it is clear that the technique could be modified to apply to other types of decision-making situations. It might therefore have applicability to various courses in the field of scientific documentation.

In conclusion, the evaluation of learning acquired is worth doing, and seems to have some relevance to the evaluation of programmes in this field. The situation is, however, somewhat complicated by the fact that we are not dealing in a subject area in which standardized tests exist or, indeed, in an area in which easily measurable skills are dealt with. It is relatively easy to measure such skills as typing ability and proficiency in translating from one language to another. It is much more difficult to test for learning acquired in an area related to the transfer of scientific and technical information. But we must not ignore this aspect of evaluation simply because it presents some difficulties.

VII. EVALUATION OF BEHAVIOURAL CHANGE

If evaluation of learning is difficult, evaluation of behavioural change in students is even more of a problem. In this phase of evaluation we are concerned with the long-term effects of an educational programme. This goes beyond learning as such into the application of the learning acquired from a particular course. The obvious consideration in this evaluation is to find out how the students have benefited, in the long run, from some programme. Benefits would most likely relate to the professional advancement of the student: promotion, increased responsibility, successful movement into a different type of position (e.g., into teaching or research), and so on. Hampton (1) suggests that one way of accomplishing the evaluation of behavioural change is by the "systematic appraisal of on-the-job performance before and after the educational experience" and that such an appraisal should be made "by the person receiving the education, his superiors, his peers, and his subordinates".

It seems quite likely that organizers of courses in this field will be interested in this level of evaluation for some at least of their educational offerings. The Unesco course conducted at Katowice in 1974, for example, had as a short-term objective the training of teachers, and prospective teachers, in the subject area of information retrieval. The longer-term goal was the further dissemination of the knowledge acquired in the course. This goal would be achieved if the course participants were able to use some or all of the course material in their own teaching, or in the planning of curricula, when they returned to their own countries. It is clear that the long-term success of this course could only be measured by some type of follow-up study, conducted perhaps six months to a year after the end of the course. A follow-up questionnaire can be used in this situation to determine whether or not the course participants have been able to make use of the material presented in their own teaching and, if they have used it, with what degree of success. At the same time, the participants can be asked to look back on the course and to assess its value within the longer-term perspective.

With the kinds of individuals and situations with which we are dealing in this particular subject field, it may not be appropriate to attempt formal on-the-job performance assessments, using peers or superiors of those participating in the programmes. Nevertheless, there may be situations in which it will be possible and appropriate to conduct some form of follow-up to gain a further and different view of the results of a programme. For example, some agency of a national government may nominate or sponsor one or more students to participate in a particular course. It may be desirable and feasible to ask this agency, some months after the completion of the course, to provide feedback on the progress and performance of these individuals. In this way, we can get some idea of how satisfied the various national governments are with the programmes we are developing and offering in this important subject field.

For most courses, workshops and seminars, it will be worth while to at least conduct a follow-up with the students. This follow-up, by means of a questionnaire, will be conducted to obtain each student's longer-term view on the value of the course, its impact on his own professional development, and the extent to which he has been able to apply the knowledge acquired.

VIII. EVALUATION OF PROGRAMME RESULTS

The "evaluation of programme results", as we are defining the expression for the purpose of this report, is concerned with the assessment of a complete educational programme, which may involve a whole series of courses or other educational endeavours. The organization responsible for planning and presenting the programme should be interested in discovering how successful the programme has been as a complete entity. This is a level of evaluation that is of broader scope than the type of evaluation we have considered so far in this report. As an example, we can evaluate as distinct units the several courses that comprise a complete curriculum in an academic department (e.g., a school of library or information science). It is conceivable that each course could be quite successful but that the entire programme fails to achieve its objectives, possibly because it is incomplete or because it emphasizes the wrong things. The programme may be successful in imparting basic technical skills and thus equipping the students to perform adequately in beginning professional positions. However, it may fail to provide the professional philosophy or theoretical underpinnings that will allow the student to "grow" in his profession and thus be capable of accepting increased managerial responsibility. If the objective of the programme is merely to instruct in basic skills it can be considered successful. If, on the other hand, the programme has the longer-range objective of equipping the student to develop professionally, to adapt to change, and to continue his learning, it may have failed completely. It is quite clear from this that we must have programme objectives as well as course objectives, and that we must develop criteria and procedures whereby the results of the programme can be evaluated against the programme objectives. Evaluation of the programme is the responsibility of the programme planner rather than the individual teachers (see the "viewpoint" categorization of Table 2), although these teachers, along with the students, may clearly have an important rôle to play in the assessment of the programme.

A series of general objectives should be developed for any programme in the area of education and training. For any specific educational activity within this programme it will be important that precise objectives, relating to the overall programme objectives, be developed. These objectives must clearly deal with who is to be trained, what information is to be imparted, and what the end results of the programmes are intended to be. The educational activities themselves must obviously be planned with these objectives clearly in mind. In relation to the stated objectives a number of important decisions must be made, decisions concerning how the objectives can be met as efficiently and economically as possible. Questions to be considered will include the following:

(1) What type of educational offering (short course, long course, workshop, seminars, or various combinations of these) is likely to be most appropriate to satisfy the programme requirements?

(2) How many of these are needed to satisfy the programme objectives and how frequently must they be presented?

(3) Where should the training programmes be held in order to have the maximum possible impact at the least cost?

(4) What individuals, types of individuals, or groups are best qualified to implement a particular programme?

(5) What type of student should be admitted to the programme, how many students should be admitted, and how should the students be selected?

(6) What language or languages should the programme be held in?

The above list is intended only as a sample of the types of questions that must be considered in the planning of training activities. The important thing to recognize here is that evaluation must be integrated into these activities from the very beginning and that it must occur at various stages in the complete process. The following sequence is recommended:

(1) establish objectives for a particular programme;

(2) evaluate the objectives. Are they really the objectives we wish to attain? Are they reasonable and viable? Modify the objectives if they fail to stand up to this examination;

(3) develop plans for a training programme that is likely to satisfy these objectives as efficiently as possible, addressing all of the various questions mentioned earlier;

(4) develop criteria by which the programme can be evaluated in terms of the achievement of its objectives;

(5) develop procedures for programme evaluation;

(6) have each element in the programme evaluated. By "element" we mean individual workshop, seminar, course, or whatever other means is used to give the education or training needed. Evaluation of a programme element will include reaction evaluation (formative and summative), evaluation of learning acquired, and evaluation of behavioural changes in the students;

(7) analyse and interpret the results of these evaluation activities. This should be a continuing and current activity. Programmes may be slightly modified, expanded, drastically changed or abandoned completely on the basis of results gathered through these evaluation activities. It is, of course, important, that the evaluation data be considered in relation to the criteria developed earlier; i.e., the criteria we identified as significant in measuring the degree to which the programme objectives have been met;

(8) on the basis of all the evaluation data available, consider whether or not the programme objectives have been satisfied, conducting whatever further surveys (e.g., follow-up studies with students, employers, sponsoring agencies) may be needed to complete the overall evaluation of programme results;

(9) identify weaknesses or failures and the causes of these. Use the knowledge thus gained in the planning of future, improved programmes.

As Steele (2) has pointed out, programme evaluation goes beyond instructional evaluation. It is concerned with the additive effects of a whole series of instructional units. Instructional evaluation is concerned with impact on the individual whereas programme evaluation is more concerned with impact on some group or community of individuals. Programme evaluation is concerned, among other things, with the establishment of priorities.

In this section of the report we have identified various broad levels of evaluation applicable to educational programmes. An example of a very thorough and multi-faceted approach to the evaluation of a training course is given in a report from the U.S. Civil Service Commission (25). In this course the behavioural objectives were carefully specified and the degree of attainment of each objective was

determined by pre- and post-testing, by post-assessment by each student and by the student's supervisor, and by "observations" made by peers and the teaching staff using a checklist. The pre- and post-test in this evaluation made use of two test sets of fifty questions each, approximately matched as to difficulty. All of these questions were of the true/false or multiple choice type. An interesting technique employed in this evaluation was the use of a self-assessment questionnaire completed by the students before and after the course. This questionnaire is designed to determine, for each element covered in the course, the student's assessment of its importance to him professionally and his own assessment of his present understanding or skill in the area. Tables 8 and 9 show the instructions presented to students and a sample page from the questionnaire. With a self-assessment instrument of this type, used before and after the course, we can determine (a) how far the course has changed the attitudes of participants towards the importance of various course elements and (b) how far the student himself feels he progressed in his knowledge of various subject areas. The authors of this report are careful to point out that, when so many approaches are taken to the evaluation of a single course, there is a very real danger that students will find the evaluation "oppressive" and will rebel against it.

The danger of "evaluation overkill" is also emphasized by Woodward and Yeager (19). If evaluation procedures are too extensive and/or redundant the evaluation process may interfere with the instruction itself. These authors also stress the need for flexibility in the approach to evaluation. We must be prepared to modify our evaluation procedures if they seem not to be giving the information we need or if we are able to identify alternative, possibly better procedures as the course progresses. This is particularly relevant, of course, to formative evaluation:

> "Formative evaluation must be 'adaptive' evaluation. This implies that the evaluation design and procedures must be sufficiently flexible to adapt to the changing requirements of the particular evaluation context in which the evaluation is being conducted, and the evaluation model must allow for such flexibility". (26)

Table 8

Instructions to Participants in a
Course on "Position Classification".
Organized by the U.S. Civil Service Commission.

SELF-ASSESSMENT QUESTIONNAIRE

This is a self-assessment questionnaire and we will ask you to complete this before and after the course. The elements listed will be covered in the course. We would like your reaction to the importance of each element for successful performance of your job and what you feel to be your present level of understanding or skill on each element.

Please indicate beside each element the importance of the element for successful performance of your job by marking in the appropriate box:

- \- if that element is of no importance or if only minor importance
- o if that element is of some importance but not essential
- \+ if that element is essential to effective performance

In addition, please circle a number value on the scale beside each element which best indicates the understanding or skill you feel you presently possess. You may circle any number on the scale. The scale is as follows:

No understanding or skill			Moderate understanding or skill					Completely adequate		
0	1	2	3	4	5	6	7	8	9	10

Understanding is evidenced by the ability to define, describe or cite the major aspects of the element.

Skill is evidenced by the ability to apply the major aspect of the element.

In addition to the questionnaire, please answer the following:

Do you presently have classification responsibilities /__/ Yes /__/ No

Understanding of the following fundamentals of employee development:

		0	1	2	3	4	5	6	7	8	9	10
1. Terminology	/__/											
2. Basic Legislation	/__/											
3. Structure of the System	/__/											
4. Methodology, Techniques and Procedures	/__/											
5. General Policy and Philosophy of the System	/__/											
6. Trends in the Field	/__/											
7. Principles	/__/											

Skill in the following:

		0	1	2	3	4	5	6	7	8	9	10
8. The Desk Audit	/__/											
9. Analysis of Position Classification	/__/											
10. Analysis of Job Information	/__/											
11. Evaluation of Positions	/__/											
12. Writing of Position Descriptions	/__/											
13. Writing of Evaluation Statements	/__/											
14. Dealing Effectively with Line Officials and Employees	/__/											

Table 9

Sample Page From Student Self-Assessment Questionnaire Used in a Course by the U.S. Civil Service Commission

IX. EVALUATING SPECIFIC EDUCATIONAL ACTIVITIES

In the earlier sections of this report we have discussed, in a general way, the subject of evaluation of education and have looked at various possible approaches and methods. This section will be concerned more specifically with the evaluation of various activities (courses, workshops, seminars) in the field of scientific documentation and will present recommendations and guidelines on how these activities might be evaluated. We assume from now on that a decision has been made that a particular educational activity will take place and that this activity is relevant to the organizer's overall objectives in the area of education and training. We are not, therefore, concerned with evaluating the need for a particular course but only with evaluating the course, in a formative and summative way, as an instructional unit and as a component in a larger programme of educational activities.

As identified by Thiede (27), five steps are involved in the evaluation of a particular activity:

(a) deciding what to evaluate,

(b) defining the behaviour desired,

(c) determining acceptable evidence,

(d) collecting evidence, and

(e) summarizing and evaluating the evidence.

We should keep all of these steps in mind throughout our planning of evaluation activities. As emphasized on several earlier occasions, it is important that consideration be given to evaluation from the very beginning of a programme. The staff of the planning body, having decided on some type of educational offering in a specific subject area, must proceed to (a) select an appropriate place and time for the meeting, and (b) select the instructor who is to have major responsibility for planning and implementing the programme. Once this is done, it will be the joint responsibility of the planning staff and of the instructor to define clearly what is to be covered (i.e., the content), who is to participate (i.e., what type of person should be attending), what the general format of the programme should be (a lecture course, a workshop, or a combination of these), and its duration. At this point the instructor must prepare an outline of the subject-matter to be covered and an indication of what level the material is to be presented at. At the same time, the organizers, working with the instructor, must draw up a list of requirements for participants in the programme (e.g., language facility, educational background, previous professional experience) and must decide how many participants will be accepted. The list of requirements will be used as the basis for selection of candidates. Participants must be selected on the basis of their previous knowledge of the subject-matter, in relation to the level set for the course. It is equally important to avoid over-qualified participants as it is to avoid the under-qualified. Some test of abilities in the language in which the course will be given seems highly desirable. The general objectives of the course must also be clearly defined at this point. That is, a statement of desired outcomes will be prepared. Anticipated outcomes must clearly be related both to the proposed content and to the type of student who will attend the programme. To take one example, the Unesco course held in Katowice in 1974 can be depicted according to the following categorization:

Subject — Information Retrieval and Information Retrieval Systems.

Level — Introductory to intermediate.

Students	Proficiency in written and spoken English. Previous experience or training in scientific documentation. Responsibility (present or future) for training information workers in their own countries.
Major objective (outcome)	To present a state-of-the art survey of modern methods of information storage and retrieval, and how this content can best be presented, so that participants will be equipped to plan or conduct similar courses in their own countries.

It will be important at this point for the instructor, together with the course organizers, to evaluate the general objectives set for the course (in relation to the other specified variables of type of instruction, type of student, location and duration of the course) to determine if they are, in fact, reasonable and realistic. The course at Katowice presented a good example of the need to evaluate stated objectives in relation to the other specified variables. Students selected to participate in this course had quite varied backgrounds. A few already had a reasonable knowledge of information storage and retrieval but the majority had little or no knowledge of the subject-matter. For a completely introductory course on the subject, some of the participants were over qualified. Without previous knowledge of the subject-matter it was unrealistic to expect this one three-week course to teach participants enough to allow them to present a similar course in their own countries. The course would, however, give enough background to allow participants, if they read and studied further at home, to develop one or more courses in this area. It should certainly have given all participants sufficient knowledge of the field that they could plan a programme in this subject area (i.e, develop a curriculum) even though they may not know enough to teach the subject themselves. On the other hand, for those with previous knowledge of the subject-matter, it would be possible to use the course as a "refresher". These students could reasonably be expected, with their previous knowledge reinforced by this course, to be able to teach the subject on return to their own countries. This particular situation can be summarized as follows:

	Student background	Emphasis of the course	Major objective or outcome
1.	No knowledge of subject-matter	Impart broad knowledge of subject-matter	Allow student to develop curricula in this area and to develop plans for training programmes
2.	Good knowledge of subject-matter already	Refresher course on subject with emphasis on content and how it should best be presented	Allow student to develop and teach specific courses in this field

Note how the pre-existing background of the student influences the direction that the course must take and the objectives that can reasonably be achieved. In the course at Katowice, where students of quite varied backgrounds participated, it was necessary to compromise. The instructors found it necessary to assume no previous background and to present a broad overview of the subject-matter. As a result, the course is likely to have achieved, for most participants at least, the first objective above but not the second.

This case illustrates clearly that, in the initial planning of a course it will be important to examine the broad objectives carefully and to choose the complete instructional staff, the mode of instruction, the duration of the course and (perhaps) its location, and, most importantly, the students in such a way that these broad objectives have the best possible chance of being achieved.

It is recommended that all educational offerings in this field make provision for both formative and summative evaluation. At the very least, the summative evaluation should be based on overall student reaction. If at all possible it would be desirable to arrive at some standard end-of-course evaluation form, a form that could be used intact (or with only minor modification) in all courses offered by a particular organization. This would allow meaningful comparisons to be made from one programme to another. In addition, if the course is designed to produce competence in a particular subject area (as opposed to dealing with "attitudinal" matters) it would be highly desirable to incorporate some method of evaluating what each student has learned. It would also be desirable to attempt to assess the long-term effects of the course; i.e., the "behavioural changes" it produces. This is likely to require a follow-up questionnaire administered to all participants some months after the course is concluded.

It is recommended that the instructor who has prime responsibility for the development of a particular course should develop a detailed outline of what the course will cover. The content must clearly be related to the overall course objectives and to the "level" of the participants. Once this outline is developed the instructor should develop, for each section of the course, one or more behavioural objectives. For each behavioural objective, the instructor must develop some demonstration of achievement. The behavioural objectives set for the Katowice course and the accompanying demonstrations of achievement (see Table 6) are put forward as an example. The more limited the objectives of the course, and the more practical its nature, the easier it will be to develop a reasonably unequivocal demonstration of achievement. For example, if a course was developed on thesaurus construction it would have one major behavioural objective (to teach students how to construct a thesaurus according to UNISIST guidelines) and one major demonstration of achievement (the ability of the student to construct a sample thesaurus in a subject field related to his own interests). If the subject coverage is quite diffuse, however, it will be necessary to have a number of behavioural objectives and a number of demonstrations of achievement as was the case with the Katowice course, which attempted a broad overview of information storage and retrieval. The less concrete or practical the subject-matter of the course the more difficult it will be to develop unequivocal demonstrations of achievement.

The specific behavioural objectives developed for a course will have an important rôle to play in its overall evaluation. If the instructor, and the organizers, consider it appropriate, these behavioural objectives can be used in the evaluation of learning acquired. That is, the students can be required to demonstrate achievement of each objective during the course itself. At the very least, however, the behavioural objectives, and stated demonstrations of achievement, should be available to all students so that, as each segment of the course is completed, they can ask themselves whether or not they feel that the behavioural objective has been satisfied and that they could in fact demonstrate its achievement appropriately. In the Katowice course the behavioural objectives were mainly used in this way although in some cases practical exercises were used as a demonstration of student achievement (e.g., exercises in indexing, the construction of searching strategies, and the development of a specimen thesaurus page).

If the behavioural objectives are not themselves to be used in the evaluation of student learning, it will be necessary to develop some other procedure for this. If the subject-matter of the course is susceptible to this kind of treatment a set of "before and after" questions can be devised. Despite the caveats presented earlier, the set of test questions used in Katowice seemed reasonably successful in providing a relatively gross indication of the extent to which learning had occurred. There was little evidence there that the participants expected to be retested and, indeed, that the lecturers placed undue emphasis on the subject areas covered by the test questions. Indeed, the score on the end-of-course test in Katowice, although

it showed a definite improvement, was far from perfect. It might be better, however, if the lecturers developed two sets of questions at approximately the same level of complexity, one to be used before the course begins and the second after it has been completed.

An alternative procedure, and perhaps the best of all, is to develop two sets of questions, A and B, and to use them in a crossover test. The group of students is also divided into two parts, one answering the A questions before the course and the B questions after it, and the other answering the B questions before the course and the A questions afterwards.

The pre-course questionnaire (see Appendix 4 for an example) should be used to determine the motivations of the students and what they hope to achieve from their participation. The questions used to test subject knowledge will be incorporated into the pre-course questionnaire. They can be used not only to measure the learning acquired by students during the course but also to give the lecturers a better understanding of what the students know at the time they enter the programme. Some modifications of the course content may need to be made on the basis of the results of a pre-course assessment of this type. If the crossover test is used we will need two pre-course questionnaires, one containing A questions and the second containing the B questions.

The formative element is perhaps the most important in the entire evaluation process. To ensure the maximum level of success the teaching staff must obtain feedback from students as the course is progressing. Although some such feedback will be obtained informally during the normal course of events, it is recommended that a formal evaluation be conducted periodically (e.g., weekly) as the course progresses. For this purpose it will be necessary to devise and apply a "reaction questionnaire". Such a form cannot be completely standardized for all courses, because each will have somewhat different features, but the major elements could remain more or less the same from one course to the next. These elements were enumerated in Table 4 and specimen questionnaires for formative evaluation purposes are illustrated in Appendix 2. The weekly reaction form used in Katowice (see Appendix 2B) was quite useful to the lecturers in determining how the course was progressing and what might later be done to improve it.

An end-of-course evaluation questionnaire will be the major component in the summative evaluation of the programme. Some of the more important questions to be answered in such a study are given in Table 5 and specimen questionnaires are shown in Appendix 3. The form used in Katowice (see Appendix 3B) was considered by the lecturers to be quite successful in measuring the overall reaction of participants to the course as a whole. The end-of-course evaluation questionnaire may also include the set of questions used to test the learning acquired by students. If the crossover test (see above) is used, we will need two such questionnaires, one containing A questions and the second B questions.

As part of the summative evaluation of the course it may be desirable to make some use of an independent observer. The observer will mix with students to obtain their reactions, observe the teaching methods used, and generally look for problems and ways in which the course might be improved. The independent observer should prepare his own report on the course. This report will supplement or complement the report prepared by the senior lecturer. It is obvious that the observer needs to be knowledgeable in the subject-matter of the course. He must also be sociable, able to talk to participants in an easy way and to encourage them to speak freely. The observer should preferably attend the entire course. If this is impractical, in the case of a long course he or she should attend selected days throughout the entire course, including the first two days and the last two days. It is not enough to have the observer appear at the end of the course. Without knowing what went on

earlier he will not get a complete picture, and the value of his contribution will be correspondingly reduced. Indeed, his interpretation and assessment of the course may be invalid if he has been unable to witness the entire programme as a logical sequence of events.

The final element in a complete evaluation will be an attempt to assess the impact of the course on those who attended; i.e., to assess its behavioural effects. This will involve at least a follow-up questionnaire mailed to all participants some months after the course is completed. Perhaps a full calendar year should elapse before such a questionnaire is administered. This questionnaire will be used to determine whether or not the students have made any use of the learning they acquired. Have the students done further work on their own in the subject-matter of the course? Have they incorporated any of the material into their own teaching? Have they developed courses or workshops themselves in the subject area of the course? Has the course had any influence on their own planning of curricula or training programmes? If the course was directed to practitioners in the field, has the course had any impact on the way they manage their own activities?

It will probably be impossible to standardize such a questionnaire because each course or workshop will have somewhat different long-term objectives. As an example, however, the follow-up questionnaire developed for use with the participants at Katowice is shown in Appendix 6.

Of course, it is not enough to simply collect evaluative data. Some use must be made of it, hopefully to improve educational programmes in the future. For each course or workshop it will be the responsibility of the teaching staff to react to the formative elements in the evaluation. They, and the course itself, must be sufficiently adaptable that the instructors are willing and able to make changes as they go along. Once the course is completed, it will be the responsibility of the senior lecturer to prepare a comprehensive report on the evaluation that has been conducted. This report must summarize the evaluation data clearly, preferably with as much as possible in tabular form, it must interpret the results as far as possible (in particular any results that can be regarded as unexpected or anomalous), and it must include the lecturer's suggestions as to how, in retrospect, the course might have been improved and his recommendations on changes that should be made in future courses of this type. An example of a final report of this kind from a data processing clinic at the University of Illinois, is included as Appendix 5.

In summary, the complete evaluation of a course, workshop or seminar should involve the following elements:

1. Close co-operation between the senior lecturer and the organizers in evaluating the objectives set for the course and in defining the major parameters of the course (size and composition of the teaching staff, size and composition of the student body, course content, duration and location) in such a way that there is a reasonable expectation that the course objectives can be met.

2. The senior lecturer must develop appropriate questionnaires, along the lines suggested in this report, for the formative evaluation of the course. Formative evaluation should take place in all educational offerings unless they are of such short duration (e.g., three days or less) that a formative element seems impractical.

3. The senior lecturer should consult with the organizers to determine the specific methods that will be used to assess the learning acquired during the course. The lecturer must develop specific behavioural objectives for the course and for each of these he must prepare one or more demonstrations of achievement. If considered appropriate by the organizers, and if the duration of the course permits it,

these demonstrations of achievement can be used in a formal way throughout the course to assess the learning taking place. If this approach is not followed, the lecturer should develop one or more sets of test questions that can be used, before and after the course, to get some indication of how much learning has actually occurred. Test questions of this kind (which must also relate to behavioural objectives) are likely to be quite effective for courses in which specific subject knowledge is to be imparted. For courses or workshops that are intended to produce "attitudinal" changes rather than to impart specific subject knowledge, the behavioural objectives and appropriate demonstrations of achievement must be used to judge how successful the course has been.*

4. Wherever possible it would be desirable to have an independent observer present at the course. The respective rôles of the teaching staff and the observer, in the total evaluation process, must be carefully defined. The independent observer should be present for the entire course or, if this is not possible, for selected days throughout the course. The independent observer should work closely with the teaching staff and he or she must have a clear understanding of the course objectives and how the lecturers are attempting to achieve these.

5. The senior lecturer should have the responsibility for designing pre-course and post-course questionnaires. The former will be used to determine the motivations of students and what they hope to gain from the course. The latter will be the major instrument used in the summative evaluation of the course. If test questions are used to assess the learning achieved, these questions will be incorporated into the pre-course and post-course questionnaires.

6. The senior lecturer will have the responsibility for designing a follow-up questionnaire to examine behavioural changes in students. This questionnaire, which is probably best designed after the course is completed, should not be sent out until 12 months have elapsed from the conclusion of the course.

7. The senior lecturer will prepare a comprehensive report on the results of his evaluation. This report will present the evaluation data together with the lecturer's interpretation and his suggestions or recommendations to the organizers. The report of the independent observer may be included as part of the lecturer's report. The lecturer should certainly get an advance copy of the observer's account and should be allowed to comment on it in his own report.

8. The preparation of the final report will be the last task required of the lecturer in the fulfilment of his contractual obligations. However, the lecturer may be asked to participate further in the follow-up study on behavioural effects and to prepare a supplementary report on the results of this survey.

---

* It is possible to develop a test to measure attitudinal changes as a result of taking a particular course. Mehlinger and Patrick (22), for example, used a Likert-scale test to measure the political attitudes of students who had taken a course on American Political Behaviour and to compare these attitudes with those of a control group.

X. DESIGN OF EVALUATION INSTRUMENTS

It has been pointed out by a number of writers in this field that the evaluation of educational programmes is not "research", at least in the strict sense of that term. In fact, as Stufflebeam (28) has indicated, there is considerable controversy as to whether experimental design has any real relevance to educational evaluation. He points out that two polar positions exist: the belief that experimental design has no relevance to the evaluation of education and the belief that educational evaluations are valid only if they are based on careful experimental design. Stufflebeam himself believes that experimental design has only a very limited application in the evaluation of education. We believe, however, that a knowledge of research methods and of principles of experimental design do have value in planning and executing even a simple evaluation programme. These subjects are outside the scope of the present report. A useful account of experimental design and the statistical analysis of survey data, as applied to the field of education, is given in a recent book by Isaac and Michael (29). Another, somewhat less detailed, is provided by Dizney (30). An excellent description of the design of educational tests in terms of their satistical validity and reliability is given in Evaluation Guidelines (31), a publication of the Nevada State Department of Education. The subject is treated in numerous other books on the evaluation of educational programmes.

Of particular relevance to the type of evaluation discussed in this report is the design of effective questionnaires. Substantial volumes have been written on the subject of questionnaire design. Books by Oppenheim (32) and by Payne (33) are particularly recommended. A very concise summary is given in a recent report by Whitney and Brandenburg (34). Although we cannot go into great detail here on the subject of questionnaire design it certainly seems desirable to mention some general principles.

It is, of course, possible to distinguish two major types of questions that may be used in an evaluation instrument: the open-ended question and the closed-ended question. In the former the respondent answers freely in narrative style while in the latter the respondent chooses an answer from a list of alternatives supplied by the investigator. The division is not an absolute dichotomy, however, for some closed-ended questions may require an open-ended explanation, as in the following example:

Have films and other audio-visual materials been used effectively throughout this course?

YES ☐

NO ☐ Please explain:

In this case the respondent is asked for more than his opinion. If he is critical of the use of audio-visual materials in the course he must explain the reasons for his criticism and/or suggest ways in which these materials might have been used more effectively. Early in this report we stressed the importance of micro-evaluation. Micro-evaluation is an analytical or diagnostic level of evaluation in which we try to discover not only what is happening but why it is happening and what might be done to improve a situation that may not now be satisfactory. It would be important to discover that most students participating in a particular course were dissatisfied with the way that audio-visual materials were used. But this in itself only indicates a specific area of dissatisfaction. It does not precisely identify the reason for the dissatisfaction. We need to know the underlying causes of dissatisfaction if we are to be able to identify an appropriate corrective action. For this reason it seems likely that we will want to make some use of open-ended questions or at

least open-ended "elaborations" on certain closed responses. Open-ended questions have a number of other advantages, including the fact that they can produce a wide variety of responses, some of which may be completely unanticipated by the investigator, that they are largely free from the influence of the investigator (it is difficult to deliberately bias the responses to an open-ended question in a printed questionnaire), and that they give respondents a chance to express their own opinions in any way they want.

The disadvantages of open-ended questions are also the advantages of questions of the closed-ended type. The latter produce greater uniformity of response, provide results that are easier to tabulate and perhaps to interpret, reduce or eliminate problems of terminology and definition, and, perhaps most important of all, reduce the amount of effort that the respondent must put into completion of the questionnaire. In a largely closed-ended questionnaire we can ask more questions and we can perhaps expect a higher rate of response. In questionnaires for course evaluation it will probably be necessary to use a mixture of closed-ended and open-ended questions. It will be desirable to use questions of the closed-ended type wherever possible, restricting the open-ended questions to situations in which it is important to get the respondent's own interpretation or his suggestions on how a situation might be improved.

Questionnaires are designed to gather two types of data: factual data and opinions or attitudes. Both types of data might be collected in the evaluation of a course or workshop. In the evaluation of learning we are likely to be primarily concerned with factual questions and factual responses. However, in other aspects of the formative and summative evaluation of a course the main emphasis will be on the attitudes of students. It is important, therefore, that we give some consideration to the subject of attitude scaling.

The simplest scale is, of course, a dichotomous one, a scale in which the respondent simply chooses between two alternatives (e.g., YES/NO or AGREE/DISAGREE). The dichotomous scale gives the respondent least flexibility in his replies. It is suitable only for questions in which, for the majority of respondents at least, the answer will be clear cut. Frequently a question of this type can be used to lead in to other questions. That is, a yes/no or agree/disagree question may be used as a screening question. For example:

Did you attend any of the sessions on thesaurus construction? NO ☐

YES ☐

If the respondent answers affirmatively he is asked some further questions about his participation in these sessions. If he answers negatively he is led on to a different section of the questionnaire.

The dichotomous scale should not be used if it seems likely that many respondents may want to qualify their responses ("Yes, but ..." or "Yes, except ..."). In this situation it is better to go to some more flexible response scale.

Multiple choice questions fall between the two extremes of the completely open-ended question and the dichotomous question. The multiple choice question is particularly suitable for use in determining attitudes or opinions because it provides for gradations in these attitudes. There are several possible scales that can be used with multiple choice questions.

Multiple choice questions involve the use of checklists from which the respondent makes his choice. We can identify three broad types of checklist, as follows:

1. the list from which the respondent chooses the single best or most appropriate response;

2. the list from which the respondent chooses all responses that apply to his particular situation;

3. the list that the respondent is asked to put in order of preference; that is, to produce a ranking of responses.

Whichever of these approaches is used it will sometimes be necessary to include an "other response" category to take care of situations that were not anticipated in advance by the designer of the questionnaire.

It is in the arrangement of multiple choice checklists that we have a number of possible scales open to us. The simplest of these scales is one in which the respondent selects a single answer from a list of possibilities as in the following example:

Select the statement that best indicates the novelty of the material presented in this course.

Most of the material was new to me ▱

About half was new to me ▱

Less than half was new to me ▱

None of the material was new to me ▱

Under certain circumstances we may want to organize our checklist in the form of a grid or matrix. This is particularly appropriate in a situation in which the same set of responses apply to several different situations, as in the following example:

For each of the instructional methods mentioned below check the box that best describes your reaction to its value as a learning experience

	Great Value	Some Value	Little Value	No Value
Transparencies				
Slides				
Films				
On-line demonstration				
Practical exercises				
Discussion periods				

A popular form of attitude scale is the so-called "Likert scale", which is quite appropriate to many applications in the evaluation of courses or workshops. In this type of scale the respondent indicates his degree of agreement with some positive or negative statement. The scale usually extends from "Strongly Agree" to "Strongly Disagree". The scale may consist of only four categories (Strongly Agree, Agree, Disagree, and Strongly Disagree) or it may include a fifth category, Undecided, that appears at the centre of the scale. It is probably more common now to omit the Undecided category thus forcing a choice among the four positive/negative response categories. In some cases the Likert scale may be expanded to seven response categories. A Likert-type scale can also be used with Like-Dislike or Accept-Reject response ranges.

The Likert scale may be applied to a complete set of statements about a course or about an instructor. About half of these statements should be favourable and about half unfavourable. The scoring of the negative statements is the reverse of the scoring of the positive statements. If we assume a four-point scale we could score the positive and negative statements according to the following scheme:

	SA	A	D	SD
Positive statement	4	3	2	1
Negative statement	1	2	3	4

In this scoring method the most positive attitude is represented by the highest score. When all the scores for all responses are averaged, those scores that are closest to four are the ones that represent the most positive attitude.

In Figures 1 and 2 are reproduced the Course Evaluation Questionnaire (CEQ) which is the form in official use by the University of Illinois at Champaign-Urbana. Notice that the form incorporates both open-ended and closed-ended questions. There are 23 closed-ended questions, some involving positive statements and some negative statements, all presented on the Likert scale. The closed-ended responses are all coded for easy data reduction by computer. The data collected in the form of responses are manipulated by statistical programmes that calculate means, standard deviations and other values. Table 10 presents a sample of the data collected in the evaluation of a particular course. The responses to each question are tabulated and the questions are then grouped into broad categories (attitude of instructor, method of instruction, and so on). Finally, a total score for the course is derived The maximum possible score for any question (and for any category), as well as the maximum possible total score is 4.00. The course represented on Table 10 scored 3.34 overall on the 4 point scale.

Another popular scale, although more complex, is the Thurstone scale. To implement this scale we must construct a series of statements about the facet of the course about which we want to obtain student attitudes. Some of these statements should be favourable to the course, some unfavourable, and some neutral. The Thurstone scale needs to be "calibrated" before it can be applied in the actual survey. This is achieved by having the set of statements pretested by a group of individuals who are "similar" to those who will participate in the actual evaluation. Each of these "judges" ranks the statements into, say, ten levels according to the degree to which the statement is considered positive or negative. For each of the statements, the distribution of the scaled values assigned by the various judges is plotted. The median of this distribution for each statement is then located. The more "ambiguous" statements (as indicated by major discrepancies among the judges as to where the statement appears on the favourable-unfavourable scale) are eliminated. From the remaining statements a set (perhaps about twenty) is selected such

Figure 1

Course Evaluation Questionnaire as
used at the University of Illinois.
Verso: closed-ended questions.
Reproduced by permission of the
University of Illinois.

FORM 73 ILLINOIS COURSE EVALUATION QUESTIONNAIRE CEQ

Measurement of Student Response by Digitek, Office of Instructional Resources

© The Board of Trustees of The University of Illinois, 1972

MARK:

- SA IF YOU STRONGLY AGREE WITH THE ITEM
- A IF YOU AGREE MODERATELY WITH THE ITEM
- D IF YOU DISAGREE MODERATELY WITH THE ITEM
- SD IF YOU STRONGLY DISAGREE WITH THE ITEM

MARK ONLY ONE RESPONSE PER ITEM USING PENCIL ONLY.

ERASE CHANGED ANSWERS CLEANLY AND COMPLETELY.

SAMPLE MARKS a) b)

1 COURSE CODE

2 ARE YOU (Number) (Name)

3 YOUR EXPECTED GRADE IN THIS COURSE

4 ARE YOU TAKING THIS COURSE FOR (PASS FAIL)

5 SPECIAL CODE

6 THIS COURSE IS WITHIN YOUR MAJOR

7 ARE YOU

8 SEMESTER

RATE THE FOLLOWING COURSE CONTENT / COURSE IN GENERAL / INSTRUCTOR: EXCELLENT, VERY GOOD, GOOD, FAIR, POOR, VERY POOR

PLEASE PRINT: THE NAME AND NUMBER OF THE MAJOR INSTRUCTOR OF THIS COURSE IS

	I would take another course that was taught this way.					BEGIN OPTIONAL ITEM SECTION I	24	SA	A	D	SD		47	SA	A	D	SD
	The instructor seemed to be interested in students as persons.						25	SA	A	D	SD		48	SA	A	D	SD
	I would have preferred another method of teaching in this course.						26	SA	A	D	SD		49	SA	A	D	SD
	It was easy to remain attentive.						27	SA	A	D	SD		50	SA	A	D	SD
	The instructor did NOT synthesize, integrate or summarize effectively.						28	SA	A	D	SD		51	SA	A	D	SD
	NOT much was gained by taking this course.						29	SA	A	D	SD		52	SA	A	D	SD
	The instructor encouraged the development of new viewpoints and appreciations.						30	SA	A	D	SD		53	SA	A	D	SD
	I learn more when other teaching methods are used.						31	SA	A	D	SD		54	SA	A	D	SD
	The course material seemed worthwhile.						32	SA	A	D	SD		55	SA	A	D	SD
	The instructor was excellent.						33	SA	A	D	SD		56	SA	A	D	SD
	The instructor demonstrated a thorough knowledge of the subject matter.						34	SA	A	D	SD		57	SA	A	D	SD
	I would rather NOT take another course from this instructor.						35	SA	A	D	SD		58	SA	A	D	SD
	It was a very worthwhile course.						36	SA	A	D	SD		59	SA	A	D	SD
14	Some things were NOT explained very well.						37	SA	A	D	SD		60	SA	A	D	SD
15	The course material was too difficult.						38	SA	A	D	SD		61	SA	A	D	SD
16	This was one of my poorest courses.						39	SA	A	D	SD		62	SA	A	D	SD
	The instructor seemed to consider teaching as a chore or routine activity.						40	SA	A	D	SD		63	SA	A	D	SD
18	It was quite interesting.						41	SA	A	D	SD		64	SA	A	D	SD
19	I think that the course was taught quite well.						42	SA	A	D	SD		65	SA	A	D	SD
20	The course content was excellent.						43	SA	A	D	SD		66	SA	A	D	SD
21	Some days I was NOT very interested in this course.						44	SA	A	D	SD		67	SA	A	D	SD
22	It was quite boring.						45	SA	A	D	SD		68	SA	A	D	SD
23	Overall, the course was good.						46	SA	A	D	SD		69	SA	A	D	SD

OPTIONAL ITEM SECTION II

70 A B C D E 71 A B C D E 72 A B C D E 73 A B C D E 74 A B C D E 75 A B C D E 76 A B C D E 77 A B C D E

78 A B C D E 79 A B C D E 80 A B C D E 81 A B C D E 82 A B C D E 83 A B C D E 84 A B C D E 85 A B C D E

DC [illegible] OPTICAL SCANNING CORPORATION

PLEASE FILL OUT THE OTHER SIDE

Figure 2

Course Evaluation Questionnaire as used at the University of Illinois. Recto: open-ended questions. Reproduced by permission of the University of Illinois.

C E Q - 51 -

PLEASE USE THIS SIDE OF THE FORM FOR YOUR PERSONAL COMMENTS ON TEACHER EFFECTIVENESS AND GENERAL COURSE VALUE--THEN TURN IT OVER AND ANSWER THE OBJECTIVE QUESTIONS ON THE OTHER SIDE, USING PENCIL ONLY. YOUR INSTRUCTOR WILL NOT SEE YOUR COMPLETED EVALUATION UNTIL AFTER FINAL GRADES ARE IN FOR YOUR COURSE.

COURSE CONTENT

PLEASE GIVE YOUR COMMENTS ON THE COURSE CONTENT, SUBJECT MATTER, AND ANY PARTICULAR RELEVANCE THIS COURSE HAS HAD TO YOUR AREA OF STUDY.

INSTRUCTORS

WRITE THE NAME OF YOUR PRINCIPAL INSTRUCTOR _____ T.A. _____

WHAT ARE YOUR GENERAL COMMENTS ABOUT THE INSTRUCTOR(S) IN THIS COURSE?

INSTRUCTIONAL OBJECTIVES

WERE THE INSTRUCTIONAL OBJECTIVES CLEARLY STATED FOR THIS COURSE? YES _____ NO _____ COMMENT:

PAPERS AND HOMEWORK

COMMENT ON THE VALUE OF BOOKS, HOMEWORK, AND PAPERS (IF ANY) IN THIS COURSE.

EXAMS

COMMENT ON THE EXAMS (QUIZZES, PRACTICALS) AS TO DIFFICULTY, FAIRNESS, ETC.

GENERAL

1. WHAT IMPROVEMENTS IN THIS COURSE WOULD YOU SUGGEST?

2. PLEASE GIVE YOUR THOUGHTFUL EVALUATION OF THIS COURSE WITH COMMENTS. ARE YOU SATISFIED WITH WHAT YOU GOT OUT OF THIS COURSE? DO YOU CONSIDER IT A VALUABLE EDUCATIONAL EXPERIENCE? SIMPLY A MEANS OF PASSING A REQUIREMENT? OR A DISAPPOINTMENT? PLEASE COMMENT.

PLEASE FILL OUT THE OTHER SIDE

U of I 31-153

OPTICAL SCANNING FORMS by OPTICAL SCANNING CORPORATION

ITEM	SA	A	D	SD	OMIT	BEST	MEAN	S.D.	DECL	0123456789
1.	0.41	0.46	0.05	0.00	0.08	SA	3.39	0.60	7	*
2.	0.72	0.26	0.00	0.03	0.00	SA	3.67	0.62	7	*
3.	0.00	0.15	0.51	0.28	0.05	SD	3.14	0.67	7	*
4.	0.26	0.51	0.15	0.05	0.03	SA	3.00	0.81	6	*
5.	0.00	0.00	0.38	0.62	0.00	SD	3.62	0.49	9	*
6.	0.00	0.03	0.41	0.51	0.05	SD	3.51	0.56	7	*
7.	0.49	0.44	0.08	0.00	0.00	SA	3.41	0.64	8	*
8.	0.00	0.08	0.51	0.33	0.08	SD	3.28	0.61	8	*
9.	0.44	0.46	0.03	0.00	0.08	SA	3.44	0.56	7	*
10.	0.49	0.46	0.03	0.03	0.00	SA	3.41	0.68	6	*
11.	0.85	0.15	0.00	0.00	0.00	SA	3.85	0.37	8	*
12.	0.05	0.08	0.36	0.51	0.00	SD	3.33	0.84	6	*
13.	0.41	0.49	0.05	0.00	0.05	SA	3.38	0.59	7	*
14.	0.00	0.21	0.49	0.26	0.05	SD	3.05	0.70	8	*
15.	0.03	0.00	0.54	0.36	0.08	SD	3.33	0.63	7	*
16.	0.03	0.03	0.26	0.62	0.08	SD	3.58	0.69	8	*
17.	0.03	0.03	0.23	0.72	0.00	SD	3.64	0.67	8	*
18.	0.26	0.51	0.13	0.00	0.10	SA	3.14	0.65	6	*
19.	0.38	0.54	0.05	0.00	0.03	SA	3.34	0.58	7	*
20.	0.36	0.41	0.13	0.00	0.10	SA	3.26	0.70	8	*
21.	0.13	0.59	0.10	0.08	0.10	SD	2.14	0.77	3	*
22.	0.05	0.03	0.33	0.49	0.10	SD	3.40	0.81	7	*
23.	0.46	0.36	0.08	0.00	0.10	SA	3.43	0.65	7	*

Explanation SA = Strongly Agree, A = Agree, D = Disagree
SD = Strongly Disagree
Omit = non-responses. Mean = average of all student responses to this question
S.D. = standard deviation of the mean
DECL = decentile in which the instructor falls (compared with all instructors in the University)
The 0 - 9 scale shows graphically where the instructor falls in terms of decentiles in relation to all other instructors
RESP = response rate for questions in this group

--SUBSCORE--	ITEMS	RESP	MEAN	S.D.	REL	RANK	LEVEL	INSTI	OVER-ALL
GENERAL ATTITUDE	4	0.93	3.48	0.50	0.81	7	7	8	8
METHOD	4	0.94	3.29	0.48	0.72	8	7	8	8
CONTENT	4	0.92	3.27	0.43	0.59	8	8	8	8
INTEREST	4	0.92	2.92	0.62	0.80	5	4	5	5
INSTR:GENERAL	2	1.00	3.37	0.68	0.74	6	6	6	6
INSTR:SPECIFIC	5	1.00	3.64	0.38	0.66	9	9	9	9
TOTAL	23	0.95	3.34	0.42	0.92	8	7	8	8

Table 10

Sample of Data from a Course Evaluation Questionnaire (see Figures 1 & 2) Completed by 39 Students

that the set contains statements that range all over the ten-point scale. The scale value for each statement selected is the mean of the rankings of the judges. The statements are put in random sequence on the questionnaire and the respondent simply checks off the statements he agrees with. A respondent's score is the sum of the scale values for the statements he has checked, the higher the score the more positive the response. The Thurstone scale is useful in determining the attitudes of individual respondents and perhaps to determine whether the attitudes have changed as a result of some additional experience. We could, for example, use the Thurstone scale to measure some attitude among students before and after taking a particular course.

A third type of scale is the so-called semantic differential. This scale involves the use of bipolar adjectives (e.g., weak/strong, good/bad, important/unimportant). The respondent is asked to rate a particular activity by marking a point on a scale between the two extremes represented by the bipolar adjectives, as in the following:

good _____ _____ _____ _____ _____ _____ _____ bad

A scale in which a respondent marks some point on a continuum is frequently referred to as a "graphic scale". A whole group of such graphic scales is created in order to measure attitude towards a particular event or activity. The semantic differential is readily converted to a numerical score reflecting the degree of "positiveness" in the attitude of the respondent, as in the examples:

weak 1 2 3 4 5 6 7 strong

complete 7 6 5 4 3 2 1 incomplete

The attitude of a student towards a course or an instructor can be evaluated by a semantic differential incorporating a whole series of graphic scales of this type.

Any type of scale in which the respondent is required to indicate his degree of satisfaction or approval is likely to be quite applicable in the design of questionnnaires for the evaluation of educational offerings. This technique was used several times in the summative evaluation of the Katowice course (see Appendix 3), using graphic scales from "Completely" to "Very incompletely", "Extremely valuable" to "Of no value at all" and "Very good" to "Very poor".

In all of the scales discussed above, the respondent checks off the single most appropriate response or he marks a particular point on a scale of values. It is also possible to include questions in which the respondent must select all statements he agrees with. A scale widely used in the evaluation of courses is the Kropp-Verner scale (35). This scale consists of a series of eighteen statements, some positive and some negative. The respondent checks all the statements he agrees with. Use of the Kropp-Verner scale is illustrated in Question 9 of the questionnaire given in Appendix 3A and the results of the application of this scale are tabulated in the report appearing in Appendix 5.

A final scale worth mentioning is the ranking scale. In this we give a list of alternatives and ask the respondent to rank these in order of priority. Again, rankings of this type have possible utility in the evaluation of courses or workshops. We could, for example, ask a student to rank the topics covered in order of their relevance to him. Or we could ask him to rank, in order of effectiveness, several approaches to instruction that might have been used in a course.

There is no particular reason why a questionnaire should use only one scale. It is perfectly acceptable to design an evaluation instrument in which a number of different scales, each most appropriate to a certain type of question, have been incorporated.

The above discussion has largely been restricted to the subject of attitude scales. Such scales are valuable, in the formative or summative evaluation of a course, to determine the attitudes or opinions of students. They could also be used to assess the attitudes of students towards various facets of the subject-matter of the course. Thus, if the course is intended to influence the students' attitudes in some way, we can test for attitude before and after the course. The construction of attitude scales is discussed and illustrated in books by Edwards (36) and by Shaw and Wright (37).

There is another subject that deserves more attention here and that is the design of tests to measure the learning gained by the students in a course. We should be interested in this type of evaluation for courses, not to grade students but to determine how effective the educational experience has been. The goal here should be to design reasonably objective tests of the learning acquired by students.

Gronlund (38) distinguishes the following major categories of tests of learning:

A. Tests in which the student supplies the answer

   1. Essay

   2. Brief notes

   3. Fill-in-the-blanks

B. Tests in which the student selects the answer

   1. True-false

   2. Matching

   3. Multiple choice

The latter group, the selection-type tests, are more controlled and objective. The answer selected by the student is either correct or it is not. Grading of such tests is quite easy. The fill-in-the-blanks type of test also produces reasonably unequivocal responses. Essays are the most difficult form of test to grade objectively and "brief notes" (which may be regarded as "mini-essays") are almost as bad.

In evaluating learning acquired during the UNISIST course in Katowice the brief notes approach was used (see Appendix 3B). This was reasonably successful. We recommend, however, that more use should be made of selection-type tests. Not only are these more objective, and thus easier to grade, but selection-type tests require less time and thus we can afford to test more extensively over the full range of subject-matter. Moreover, selection-type tests have another great advantage in the international environment. The student whose native tongue is not that of the course will be at less of a disadvantage in this type of test. In fact, it should be relatively easy to produce such a test in several languages to reduce this disadvantage even further.

With a little ingenuity most tests of learning can be reduced to the selection format. This can be illustrated by a restructuring of some of the questions used in the Katowice evaluation. Question 13 (Appendix 3B), for example, was presented as follows:

With what developments in information retrieval do you associate the following names:

Willian Batten

Calvin Mooers

Mortimer Taube

Eugene Garfield

Hans Peter Luhn

This question could have been handled in many other ways. It is a "natural" for example for use in a matching test, as follows:

Match the names in Column A with the developments in information retrieval shown in Column B. Put the appropriate number against each name.

A	B
William Batten	1. SDI
Calvin Mooers	2. Uniterm
Mortimer Taube	3. Citation indexing
Eugene Garfield	4. Peek-a-boo
Hans Peter Luhn	5. Edge-notched cards

The question could also be handled, although somewhat more laboriously, in a true/false format:

Mortimer Taube was the first person to use peek-a-boo cards for information retrieval

TRUE

FALSE

or in a multiple-choice approach:

Mortimer Taube:

a. invented citation indexing ▱

b. developed the first SDI system ▱

c. produced the Uniterm system ▱ SELECT ONE

d. invented the peek-a-boo principle ▱

e. pioneered the use of edge-notched cards for information retrieval ▱

Virtually all factual-type tests of learning can be handled in this way. Consider Question 12 in Appendix 3B. This was designed to determine whether or not the student understood the difference between an on-line and an off-line system for

information retrieval and whether or not he recognized the advantages of the on-line system over the off-line. This could have been handled as follows:

An on-line system for information retrieval:

a. Can only be used by information specialists ☐

b. Involves a conversational interaction between the user and the system ☐

c. Permits browsing ☐

d. Does not produce a printout of search results ☐

e. Cannot be used for searches in which rapid response is needed ☐

f. Is essentially non-heuristic ☐

SELECT ALL APPLICABLE STATEMENTS

and so on.

Before we design any type of test of learning we must obviously define the learning outcomes that are to be measured in the test. In the above discussion we have been assuming a course or workshop whose prime purpose was the transfer of specific subject knowledge or skills. It is this type of educational endeavour that we are most likely to be concerned with. It should be possible to test objectively the extent to which the subject knowledge or skills are transferred. But a course or workshop may have other objectives. It may be designed, for example, to improve the problem-solving or decision-making skills of the participants. It is not inconceivable that a test of problem-solving or decision-making skills could be reduced to a selection-type exercise. It is more likely, however, that we will need to design something more elaborate. Tests of decision making or problem solving can be designed for the information science situation (e.g., problems involved in managing an information centre) along the lines of the tests of decision making in health care that were mentioned earlier in this report.

A technique that might be used in the evaluation of courses designed to produce attitudinal changes, or to improve problem-solving and decision-making skills, is the student self-evaluation. This involves a self-rating scale. Students judge their own position on this scale before and after the course or workshop. Burstiner (39), for example, used a scale of 20 "creativity" traits and 20 "leadership" traits in evaluating a workshop on creative leadership. Each participant rated himself on each of these traits on a seven-point scale ranging from "extremely low" to "extremely high". Significant gains were recorded in both creativity and leadership scales after the course was completed. An example of a self-assessment questionnaire is given in Tables 8 and 9 of this report.

More complete discussions of the design of tests of learning are given in books by Gronlund (38) (40), by Ahmann and Glock (41), Bloom et al (42), and Lindquist (43).

XI. CONCLUSION

One important matter, alluded to earlier but not elaborated on, is that of the adequacy of the evaluation information that is collected. Knox (44) identifies five facets of "adequacy" in this context:

1. Validity
2. Reliability
3. Universality
4. Application of results
5. Efficiency

The adequacy of evaluation data is also discussed by Siegel et al (45), who also provide a very useful review of various approaches to the design of achievement tests.

In terms of validity the important consideration is to ensure that the data collected do in fact tell us what we want to know. This implies that the evaluation, as well as the educational offering itself, must have defined objectives. That is, we must first carefully specify what it is we want to learn about the course and must then design evaluation procedures that will gather the data we need to draw our conclusions. We must be certain that our evaluation procedures, collectively, gather all the data we need, and equally important, that we are not collecting superfluous data for which we will have no particular use. Also, we must be sure that the questions we ask are questions that the respondents are competent to answer. If we are looking for completely frank responses, we may be more likely to get these if questionnaires are completed anonymously. Likewise, the anonymous approach may be preferable in any tests of learning acquired by students. A student may feel less threatened if he is not required to put his name on a test of this kind.

Reliability relates to the reproducibility of evaluation results. A questionnaire, or other evaluation instrument, may be considered reliable if it gives approximately the same results however many times it is applied in an identical situation. We could test the reliability of an evaluation questionnaire by applying it twice under exactly the same conditions (i.e., using the same group of students at exactly the same point in a course). This could be done, for example, by having the questionnaire completed before and after lunch on a particular day, but this does not seem a very practical or desirable approach. Usually we must be satisfied with less. At the very least, we must do everything we can to ensure that the questions we ask are quite clear and that they are not susceptible to various interpretations. Closed-ended questions (i.e., those requiring a selection from a list of alternatives) are likely to get more consistent responses than open-ended questions. Standard scales, such as the Kropp-Verner scale mentioned earlier, are likely to produce quite reliable responses because these have been tested and proven with many previous audiences. It is desirable that a questionnaire be pretested before a course by asking various people, having roughly the same background as the actual participants, for their interpretation of the meaning of the questions. A question that is always interpreted in the same way, by members of this pretest group, may be regarded as unambiguous and reliable. A question that is variously interpreted will require modification for purposes of clarification. As more courses are held by a particular organization, and more evaluations are conducted, it will be possible to identify questions or types of questions that appear to produce reliable responses. Wherever possible, it seems desirable to develop fairly standardized evaluation instruments -

questionnaires that can be applied, perhaps with minor modification, to several courses. This has two advantages: (a) reliability is improved by the use of questions that have been tested in the past, and (b) we can compare the impact of various courses if the evaluation methodology is held constant.

This brings us logically to the next characteristic, that of universality. Each educational offering within a particular programme will have some unique features that we may want to evaluate. But the various programmes will also have many things in common. If the same approach is used to the evaluation of the common elements it will be possible, as Knox (44) has pointed out, to combine findings in order to develop improved guidelines for the planning and implementation of courses. Analysis of responses to common elements would allow the planning body to identify characteristics of courses or participants that, in the words of Knox, "are associated with high satisfaction or high achievement and those associated with low satisfaction or achievement". It is strongly recommended that a sponsoring organization should encourage the development of standard approaches to the evaluation of courses, workshops and seminars. Each set of evaluation instruments should include both common elements and elements that may be needed to evaluate the unique features of a particular programme.

The characteristic application of results obviously relates to the use that is to be made of the evaluation data. We must be quite sure that we collect data in such a way that we can apply it to answer all the questions that we would like to answer about a particular programme. To take a simple example, we may want to know what effect the amount of previous professional experience has on the participants' reaction to a course or on the degree to which they seem to have benefited from it. Quite obviously, if forms are completed anonymously, we must ask respondents to indicate the number of years they have been active in the field. In this way we can separate the responses of the more experienced participants from those of less experience. This highlights once more the need to prepare evaluation specifications; that is, to define very carefully in advance what it is that we want to learn about a particular course. It seems exteemely important, then, that those responsible for planning a course should prepare a set of specific questions for which they would like to obtain answers. The evaluation should then be designed to answer these questions as reliably, thoroughly and efficiently as possible.

Efficiency relates to the way data is collected. It is the responsibility of the evaluator to see that the needed data are collected in the easiest possible way from those who are in the best position to provide it. The evaluator must avoid redundancy: he must not collect data that he will not use and he must not collect the same data twice (unless he feels he needs to as a check on reliability). Questionnaires should not be used to gather data that are readily available in other ways (e.g., the application forms of participants). The data-gathering procedures should interfere as little as possible with the conduct of educational programmes. We must avoid evaluation "overkill" at all costs. The evaluator must also consider the best time to collect various data as well as the best way of collecting it. It is also important to plan the data collection in such a way that reduction, analysis, and interpretation of the data are facilitated. Because the majority of documentation courses will involve relatively small numbers of participants, it is likely that the data reduction can be handled quite adequately by hand. If very large quantities of data are to be collected, however, serious consideration should be given to the possibility of collecting these data in such a way that they may be manipulated by standard data processing techniques. The report by Whitney and Brandenburg (38) contains useful guidelines on the design of questionnaires to facilitate subsequent data processing operations.

This report has attempted to provide guidelines that can be used in the evaluation of courses, workshops and seminars in the field of scientific and technical

information and documentation. We have deliberately restricted ourselves to the presentation of guidelines rather than trying to lay down rigid rules or "standards" for the conduct of evaluation programmes. The guidelines relate to what data to collect and how, in general, they may be collected. The guidelines are illustrated with examples of questionnaires, tests and reports. This is not, however, a cook-book approach. We are not presenting recipes to be slavishly followed. On the other hand, we have tried to present various ingredients from which a recipe for a particular evaluation could be put together. Although, as we have stated several times already, it is desirable to standardize the collection of data representing elements common to several courses, there is also considerable room for experimentation and innovation in the evaluation of individual offerings within a particular programme. Such experimentation and innovation should be encouraged.

An encyclopaedia of terms and concepts related to the evaluation of education has been compiled by Anderson et al (46). This volume may serve as a useful reference tool for all involved in the planning and implementation of evaluation activities.

XII. REFERENCES

1. Hampton, L.A. Evaluating continuing education programs. Adult leadership (Mount Morris, Illinois), vol. 22, no. 3, September 1973, pp. 105-119.

2. Steele, S.M. Contemporary approaches to program evaluation and their implications for evaluating programs for disadvantaged adults. Syracuse, N.Y., ERIC Clearinghouse on Adult Education, 1973.

3. Bergman, B.A.; Siegel, A.I. Training evaluation and student achievement measurement: a review of the literature. Wayne, Pa., Applied Psychological Services, Inc., January 1972. 67 p. ED 066893.

4. Knox, A.B. Approaches to conference evaluation. Working paper prepared for the Conference and Institute Division Meeting, National University Extension Association, Lincoln, Nebraska, 29 April-2 May 1962. Available from the Office for the Study of Continuing Professional Education, University of Illinois, Urbana, Ill.

5. Scriven, M.S. The methodology of evaluation. In: R.W. Tyler et al (eds.). Perspectives of curriculum evaluation, pp. 39-83. Chicago, Rand McNally, 1967. 102 p., illus. (AERA Monograph Series on Curriculum Evaluation, 1).

6. Steele, S.M. Contemporary approaches to program evaluation, p. 131.

7. Ibid., p. 7.

8. Ibid., p. 236.

9. Kirkpatrick, D.L. Evaluation of training. In: Craig, R.L.; Bittel, L.R. (eds.). Training and development handbook, pp. 87-112. New York, McGraw-Hill, 1967, xii, 650 p., illus.

10. Knox, A.B. Program evaluation in adult basic education. Tallahassee, Florida State University, 1971.

11. Stake, R.E. The countenance of educational evaluation. Teachers College record (New York), vol. 68, no. 7 April 1967, pp. 523-540.

12. Stake, R.E. A research rationale for EPIE. The EPIE forum (New York), vol. 1, no. 1, September 1967, pp. 7-15.

13. Stake, R.E. Objectives, priorities and other judgment data. Review of educational research (Washington, D.C.), vol. 40, no. 2, April 1970, pp. 181-212.

14. Roid, G.H. Systems design for course evaluation. Paper presented at the AERA Symposium "Research, Systems Design and the Future of University Course Evaluation", Chicago, April 1972. 43 p. ED 064 398.

15. Provus, M. Evaluation of ongoing programs in the public school system. In: Tyler, R.W. (ed.). Educational evaluation: new roles, new means, pp. 242-283. Chicago, University of Chicago Press, 1969. xiii, 409 p. (68th Yearbook of the National Society for the Study of Education, part 2.).

16. Stake, R.E. Responsive evaluation. December 1972. 4 p. ED 075 487.

17. Knowles, M.S. The modern practice of adult education. New York, Association Press, 1970. p. 224. 384 p., illus.

18. Hampton, L.A. Evaluating continuing education programs. p. 107.

19. Woodward, J.P.; Yeager, J.L. Evaluation of programs to train educational R & D personnel. Pittsburgh, University of Pittsburgh, Learning Research and Development Center, April 1972. 26 p. ED 064 376.

20. Knowles, M.S. The modern practice of adult education, p. 237.

21. Hunter, M.G.; Schooley, D.E. The synergistic evaluation model. 1973. 10 p; ED 075 503.

22. Mehlinger, H.D.; Patrick, J.J. The use of "formative" and "summative" evaluation in an experimental curriculum project: a case in the practice of instructional materials evaluation. Bloomington, Indiana University, High School Curriculum Center in Government, March 1970. 16 p. ED 041 443.

23. Veri, C.C. et al Evaluating teaching effectiveness in extension-administered programs. St. Louis, University of Missouri, Extension Division, 1973. 15 p. ED 078 275.

24. McGuire, C.H.; Babbott, D. Simulation technique in the measurement of problem solving skills. Journal of educational measurement (East Lansing, Michigan), vol. 4, no. 1, Spring 1967, pp.1-10.

25. U.S. Civil Service Commission. An evaluation model designed to measure the effectiveness of a one-week training course "Position Classification and the Management Process". Washington, D.C., June 1970. 48 p. ED 064 468.

26. Woodward, J.P.; Yeager, J.L. Evaluation of programs to train educational R & D personnel. p. 13.

27. Thiede, W. Evaluation and adult education. In: Jensen, G. et al (Eds.). Adult education: outlines of an emerging field of university study, pp. 291-305. Washington, D.C., Adult Education Association of the U.S.A., 1964, xiv, 334 p.

28. Stufflebeam, D.L. The use of experimental design in educational evaluation. Journal of educational measurement (East Lansing, Michigan), vol. 8, no. 4, Winter 1971. pp. 267-274.

29. Isaac, S.; Michael, W.B. Handbook in research and evaluation. San Diego, R.R. Knapp, 1971, vi, 186 p., illus.

30. Dizney, H. Classroom evaluation for teachers. Dubuque, Iowa, W.C. Brown, Co., 1971. xii, 119 p., illus. (Issues and innovations in education).

31. Evaluation Guidelines. Carson City, Nevada, State Department of Education, July 1973. 75 p. ED 086 727.

32. Oppenheim, A.N. Questionnaire design and attitude measurement. New York, Basic Books, 1966, ix, 208 p., illus. (Basic topics in sociological method).

33. Payne, S.L. The art of asking questions. Princeton, Princeton University Press, 1951. xiv, 349 p., bibliogr. (Studies in public opinion, 3).

34. Whitney, D.R.; Brandenburg, D.C. The questionnaire as a data source. Urbana, Illinois, University of Illinois, Office of Instructional Resources, August 1974. 45 p.

35. Kropp, R.P.; Verner, C. An attitude scale technique for evaluating meetings, Adult education (Chicago), vol. 7, no. 4, Summer 1957, pp. 212-215.

36. Edwards, A. Techniques of attitude scale construction. New York, Appleton-Century-Crofts, 1957. 256 p., illus. (Century psychology series).

37. Shaw, M.E.; Wright, J.M. Scales for the measurement of attitudes. New York, McGraw-Hill, 1964. xxii, 604 p. (McGraw-Hill series in psychology).

38. Gronlund, N.E. Constructing achievement tests. Englewood Cliffs, New Jersey, Prentice-Hall, 1968. ix, 118 p., illus.

39. Burstiner, I. Evaluation of a mini-workshop in "Creative Leadership" for educational administrators. Education (Milwaukee, Wisconsin), vol. 93, no. 1, September-October 1972, pp. 47-50.

40. Gronlund, N.E. Measurement and evaluation in teaching. New York, Macmillan, 1965. xiii, 420 p., illus.

41. Ahmann, S.; Glock, M.D. Evaluating pupil growth; principles of tests and measurements. 4th ed. Boston, Allyn and Bacon, 1971. 621 p.

42. Bloom, B.S.; Hastings, J.T.; Madaus, G.F. Handbook on formative and summative evaluation of student learning. New York, McGraw-Hill, 1971. x, 923 p., illus.

43. Lindquist, E.F. (ed.). Educational measurement. Washington, D.C., American Council on Education, 1951. xix, 819 p., illus.

44. Knox, A.B. Approaches to conference evaluation. p. 11.

45. Siegel, A.I.; Bergman, B.A.; Federman, P. Some techniques for the evaluation of technical training courses and students. Wayne, Pa., Applied Psychological Services, Inc., February 1972. 137 p. AD 753 094.

46. Anderson, S.B.; Ball, S.; Murphy, R.T.; Rosenthal, E.J. Anatomy of evaluation: important concepts and techniques in evaluating education/training programs. Princeton, N.J., Educational Testing Service, August 1973. 412 p. AD 767 713.

APPENDICES

1. Selective Bibliography on the Evaluation of Education

2. Examples of Questionnaires Used in the Formative Evaluation of a Course or Workshop

3. Examples of Questionnaires Used in the Summative Evaluation of a Course or Workshop

4. Pre-Course Questionnaire Used in Evaluation of Learning Occurring in the UNISIST Course on Information Retrieval and Information Retrieval Systems, Katowice, Poland, August 1974

5. Report on the Evaluation of the Eleventh Annual Clinic on Library Applications of Data Processing, University of Illinois, May 1974

6. Follow-up Questionnaire Developed to Evaluate Impact of the Course Held in Katowice, Poland, August 1974

7. Example of a Checklist for Use by an Independent Observer

APPENDIX 1

Selective Bibliography on the Evaluation of Education

Adkins, D.C. Measurement in relation to the educational process. Educational and psychological measurement. (Durham, N.C.), vol. 18, no. 2, Summer 1958, pp. 221-240.

Ahmann, S.; Glock, M.D. Evaluating pupil growth: principles of tests and measurements. 4th ed. Boston, Allyn and Bacon, 1971. 621 p.

Alexander, F.D. A critique of evaluation. Journal of co-operative extension. (Madison, Wis.), vol. 3, no. 4, Winter 1965, pp. 205-212.

Alkin, M.C. Objectives and objective-based measures in evaluations. Washington, D.C., Office of Education, June 1970. 13 p. ED 043 666.

_____. Towards an evaluation model: a systems approach. Los Angeles, University of California, Center for the study of evaluation of instructional programs, August 1967. 21 p., illus., bibliogr. (Working paper no. 4.).

Angell, D.; Shearee, J.W.; Berliner, D.C. Study of training performance evaluation techniques. Port Washington, U.S. Naval Training Device Center, 1964. vi, 74 p. illus. (Technical report NAVTRADEVCEN 1449-1.).

Armstrong, R.; Cornell, T.D.; Kraner, R.; Roberson, E.W. A scheme for evaluation. In: Roberson, E.W. (ed.). Educational accountability through evaluation, pp. 19-30. Englewood Cliffs, New Jersey, Educational Technology Publications, 1971. 107 p. illus.

Bail, F.T.; Dunn-Rankin, P. The evaluation of teaching effectiveness. Educational perspectives, (Honolulu, Hawaii), vol. 11, no. 2, May 1972, pp. 26-32.

Bergman, B.A.; Siegel, A.I. Training evaluation and student achievement measurement: a review of the literature. Wayne, Pa., Applied Psychological Services, Inc., January 1972. 67 p. ED 066 893.

Bloom, B.S.; Hastings, J.T.; Madaus, G.F. Handbook on formative and summative evaluation of student learning. New York, McGraw-Hill, 1971. x, 923 p., illus.

Bloom, B.S. Toward a theory of testing which includes measurement-evaluation-assessment Los Angeles, University of California, 1968. 41 p., bibliogr. (Center for the study of evaluation of instructional programs occasional report no. 9).

Borus, M.E.; Tash, W.R. Measuring the impact of manpower programs: a primer. Ann Arbor, Michigan, the University of Michigan - Wayne State University, Institute of Labor and Industrial Relations, 1970. 81 p. (Policy papers in human resources and industrial relations 17).

Boyle, P.G.; Jahns, I.R. Program development and evaluation. In: Smith, R.M.; Aker, G.R.; Kidd, J.R. (eds.). Handbook of adult education, pp. 59-74. New York the Macmillan Company, 1970. xxix, 594 p.

Burstiner, I. Evaluation of a mini-workshop in "Creative Leadership" for educational administrators. Education (Milwaukee, Wis.), vol. 93, no. 1, September-October 1972, pp. 47-50.

Caro, F.G. (ed.). Readings in evaluation research. New York, Russell Sage Foundation, 1971. xiv, 418 p., illus., bibliogr. (Publications of the Russell Sage Foundation).

Catalanello, R.F.; Kirkpatrick, D.L. Evaluating training programs - the state of the art. Training and development journal (Montclair, N.J.), vol. 22, no. 5, May 1968, pp. 2-9.

Collins, A.I. Assessing the effectiveness of industrial training. Canberra, Australia, Commonwealth Public Service Board, Training Section, 1961.

Crane, P.; Abt, C.C. A model for curriculum evaluation. Educational technology (Saddlebrook, N.J.), vol. 9, no. 10, October 1969, pp. 17-25.

Cronbach, L.J. Course improvement through evaluation. In: Taylor, P.A.; Cowley, D.M. (eds.). Readings in curriculum evaluation, pp. 11-19. Dubuque, Iowa, Wm. C. Brown Company Publishers, 1972. x, 253 p., illus., bibliogr.

Deniston, O.I.; Rosenstock, I.M.; Welch, W.; Getting, V.A. Evaluation of program efficiency. Public health reports (Rockville, Md.), vol. 83, no. 7, July 1968, pp. 603-610.

Dizney, H. Classroom evaluation for teachers. Dubuque, Iowa, W.C. Brown Co., 1971. xii, 119 p., illus. (Issues and innovations in education).

Donnelly, F.A.; Ware, J.E.; Wolkob, G.H.; Naftulin, D.H. Evaluation of weekend seminars for physicians. Journal of medical education (Chicago), vol. 47, no. 3, March 1972, pp. 185-187,

Ebel, R.L. Measuring educational achievement. New York, Prentice-Hall, 1965. xii, 481 p., illus., bibliogr. (Prentice-Hall education series).

Educational evaluation. Review of educational research (Washington, D.C.), vol. 40, no. 2, April 1970, pp. 181-320.

Ellner, C.L. Summative evaluation of a curriculum developed to prepare day care administrators using mastery/non-mastery criteria. Claremont Graduate School, California Center for Early Education, April 1972. 36 p. ED 065 553.

Evaluation guidelines. Carson City, Nevada, State Department of Education, July 1973. 75 p. ED 086 727.

Glaser, R. Institutional technology and the measurement of learning outcomes: some questions. American psychologist (Washington, D.C.), vol. 18, no. 8, August 1963, pp. 519-521.

Glass, G.V. Educational product evaluation: a prototype format applied. Educational researcher (Washington, D.C.), vol. 1, no. 1, January 1972, pp. 7-10.

_____. The growth of evaluation methodology. Boulder, University of Colorado, 1969. 15 p., bibliogr. (Laboratory of educational research. Research papers, no. 27).

Gottman, J.M.; Clasen, R.E. Evaluation in education: a practitioner's guide. Itasca, Illinois, F.E. Peacock Pulbishers, Inc., 1972. xvii, 512 p., illus., bibliogr. (management series in education).

Gronlund, N.E. Constructing achievement tests. Englewood Cliffs, New Jersey, Prentice-Hall, 1968. ix, 118 p., illus.

_____. Measurement and evaluation in teaching. New York, Macmillan, 1965. xiii, 420 p., illus.

Grotelueschen, A.D.; Gooler, D.D. A role of evaluation in planning educational programs. February 1971. 15 p. ED 052 221.

_____. _____. Knox, A.B.; Kemmis, S.; Dowdy, I.; Brophy, K. An evaluation planner. Urbana, University of Illinois, Office for the Study of Continuing Professional Education, 1974. v, 121 p.

Hampton, L.A. Evaluating continuing education programs. Adult leadership (Mount Morris, Ill.), vol. 22, no. 3, September 1973, pp. 105-119.

Hemphill, J.K. The relationship between research and evaluation studies. In: Tyler, R.W. (ed.). Educational evaluation: new roles, new means, pp. 189-220. Chicago, the University of Chicago Press, 1969. xiii, 409 p. (68th Yearbook of the National Society for the Study of Education, Part II).

Hertzog, W.T. Costs of evaluation. In: Proceedings: evaluation in continuing medical education, pp. 26-35. the University of Kansas Medical Center, August 1970.

Hertzog, E. Some guidelines for evaluative research. Washington, D.C., U.S. Department of Health, Education, and Welfare, Social Security Administration, 1959. 117 p. (U.S. - Children's Bureau. Publication no. 375).

Hesseling, P. Strategy of evaluation research in the field of supervisory and management training. Assen, The Netherlands, Van Gorcum and Company, Ltd., 1966. 364 p. bibliogr.

Hoke, G.A. Evaluation and accountability: a resource unit for educators. Urbana, University of Illinois, Center for Instructional Research and Curriculum Evaluation, July 1971. 48 p. ED 086 728.

Hunter, M.G.; Schooley, D.C. The synergistic evaluation model. 1973. 10 p. ED 075 503.

Issac, S.; Michael, W.B. Handbook in research and evaluation. San Diego, R.R. Knapp, 1971. vi, 186 p., illus.

Kirkpatrick, D.K. How to start an objective evaluation of your training programs. Training and development journal (Montclair, N.J.), vol. 10, no. 3, May-June 1956, p. 18.

Kirkpatrick, D.L. Evaluation of training. In: Craig, R.L.; Bittel, L.R. (eds.). Training and development handbook, pp. 87-112. New York, McGraw-Hill, 1967. xii, 650 p., illus.

Knowles, M.S. The modern practice of adult education. New York, Association Press, 1970. 384 p., illus.

Knox, A.B. Conference objectives: prelude to evaluation. Adult leadership. (Mount Morris, Ill.), vol. 10, no. 8, February 1962, p. 234 ff.

_____. Continuous program evaluation. In: Shaw, N.C. (ed.). Administration of continuing education, a guide for administrators, pp. 368-391. Washington, D.C. National Association for Public School Education, 1969. x, 483 p., illus., bibliogr.

_____. Developing an evaluation plan in adult farmer education. Agricultural education magazine (Des Moines, Iowa), vol. 35, no. 12, December 1962, p. 100ff.

_____. Program evaluation in adult basic education. Tallahassee, Florida State University, 1971.

Knox, A.B.; Mezirow, J.; Darkenwald, G.G.; Beder, H. An evaluation guide for adult basic education programs. New York, Columbia University, Teachers College, Center for Adult Education, 1974. 13 p.

Kravetz, N. The evaluation of educational system outputs: an exploratory study. Paris, Unesco, International Institute for Educational Planning. 107 p. ED 055 321. (Report no. IIEP-RP-8).

Matteson, H.R. The process of evaluating vocational educational programs in agriculture. Madison, Wisconsin, University of Wisconsin, Department of Agricultural and Extension Education, February 1972. (Instructional series no. 3).

McGuire, C.H.; Babbott, D. Simulation technique in the measurement of problem solving skills. Journal of educational measurement (East Lansing, Mich.), vol. 4, no. 1 Spring 1967, pp. 1-10.

Mehlinger, H.D.; Patrick, J.J. The use of "formative" and "summative" evaluation in an experimental curriculum project: a case in the practice of instructional materials evaluation. Bloomington, Indiana University, High School Curriculum Center in Government, March 1970. 16 p. ED 041 443.

Mehrens, W.A.; Ebel, R.L. (eds.). Principles of educational and psychological measurement. Chicago, Rand McNally, 1967. ix, 483 p., illus., bibliogr. (Rand McNally education series.).

Miller, H.L.; McGuire, C.H.; Evaluating liberal adult education. Chicago, Center for the Study of Liberal Education for Adults, 1961. 184 p. (research reports.).

Morgan, B.; Holmes, G.E.; Bundy, C.E. Methods in adult education. Danville, Illinois, the Interstate Publishing Company, 1960. 180 p., illus., bibliogr.

Nakamoto, J.; Verner, C. Continuing education in the health professions: a review of the literature pertinent to North America. Syracuse, New York, ERIC Clearinghouse of Adult Education, March 1973. 340 p. ED 074 367.

Popham, W.F. An evaluation guidebook. Los Angeles, The Instructional Objectives Exchange, 1972. 89 p.

Provus, M. Discrepancy evaluation for educational program improvement and assessment. Berkeley, California, McCutchan Publishing Corp., 1971. ix, 380 p., bibliogr.

_____. Evaluation of ongoing programs in the public school system. In: Tyler, R.W. (ed.). Educational evaluation: new roles, new means, pp. 242-283. Chicago, The University of Chicago Press, 1969. xiii, 409 p. (68th Yearbook of the National Society for the Study of Education, part II.).

Roid, G.H. Systems design for course evaluation. Paper presented at the AERA Symposium "Research, Systems Design and the Future of University Course Evaluation", Chicago, April 1972. 43 p. ED 064 398. 73 mo

Schulberg, H.C.; Baker, F. Program evaluation models and the implementation of research findings. In: Schulberg, H.C.; Sheldon, A.; Baker, F. (eds.). Program evaluation in the health fields, pp. 562-572. New York, Behavioral Publications Inc., 1969. xviii, 582 p., illus., bibliogr. (Social problems series.).

Scriven, M.S. The methodology of evaluation. In: Tyler, R.W. et al (eds.). Perspectives of curriculum evaluation, p. 39-83. Chicago, Rand McNally, 1967. 102 p., illus. (AERA monograph series on curriculum evaluation, 1.).

_____. Pros and cons about goal-free evaluation. Evaluation comment (Los Angeles, Calif.), vol. 3, no. 4, December 1972, pp. 1-8.

Seaman, B.F.; Dutton, D. Self evaluation. In: Klevins, C. (ed.). Materials and methods in adult education, pp. 265-274. Canoga Park, California, Klevens Publications, Inc., 1972. 373 p., bibliogr. (Klevens adult education series.).

Smith, R.G., Jr. Controlling the quality of training. Washington, D.C., George Washington University, Human Resources Research Office, 1965. x, 41 p., illus., bibliogr. (Human Resources Research Office technical report 65-6.).

Stake, R.E. The countenance of educational evaluation. Teachers College Record (New York), vol. 68, no. 7, April 1967, pp. 523-540.

_____.; Denny, T. Needed concepts and techniques for utilizing more fully the potential of evaluation. In: Tyler, R.W. (ed.). Educational evaluation: new roles, new means, pp. 370-390. Chicago, The University of Chicago Press, 1969. xiii, 409 p. (68th Yearbook of the National Society for the Study of Education, part II.).

_____. Objectives, priorities and other judgment data. Review of educational research (Washington, D.C.), vol. 40, no. 2, April 1970, p. 181-212.

_____. A research rationale for EPIE. The EPIE forum (New York), vol. 1, no. 1, September 1967, pp. 7-15.

_____. Responsive evaluation. December 1972, 4 p. ED 075 487.

_____. Toward a technology for the evaluation of educational programs. In: Tyler, R.W.; Gagne, R.; Scriven, M. (eds.). Perspectives of curriculum evaluation, pp. 1-12. Chicago, Rand McNally, 1967. 102 p., bibliogr. (AERA monograph series on curriculum evaluation, 1.).

Steele, S.M. Contemporary approaches to program evaluation and their implications for evaluating programs for disadvantaged adults. Syracuse, N.Y., ERIC Clearinghouse on Adult Education, 1973. 266 p. ED 075 715.

_____. Cost-benefit analysis and the adult educator: a literature review. Washington, D.C., ERIC Clearinghouse on Adult Education and Adult Education Association of the U.S.A., July 1971, 34 p. ED 056 287.

_____. Program evaluation - a broader definition. Journal of extension (Madison, Wis.), vol. 8, no. 2, Summer 1970, pp. 5-17.

Stufflebeam, D.L. The relevance of the CIPP evaluation model for educational accountability. Journal of research and development in education (Athens, Ga.), vol. 5, no. 1, Fall 1971, pp. 19-25.

_____. The use of experiemntal design in educational evaluation. Journal of education measurement (East Lansing, Mich,), vol. 8, no. 4, Winter 1971, pp. 267-274.

Suchman, E.A. Evaluative research: principles and practice in public service and social action programs. New York, Russell Sage Foundation, 1967. ix, 186 p., bibliogr.

Taba, H.; Sawin, E. A proposed model of evaluation. Educational leadership (Washington, D.C.), vol. 20, no. 1, October 1962, pp. 57-59 ff.

Tamer, C.K. Program evaluation in cost-benefit terms. Paper presented at the AERA Annual Meeting, Minneapolis, Minn., 2-6 March, 1970. 15 p. ED 042 228.

Taylor, P.A.; Crowley, D.M. (eds.). Readings in curriculum evaluation. Dubuque, Iowa, Wm. C. Brown Company Pulbishers, 1972. x, 253 p., illus., bibliogr.

_____.; Maguire, T.O. A theoretical evaluation model. Manitoba journal of educational research (Fort Gerry, Manitoba), vol.1, June 1966, pp.12-17.

Thiede, W. Evaluation and adult education. In: Jensen, G. et al (eds.). Adult education: outlines of an emerging field of university study, pp.291-305. Washington, D.C., Adult Education Association of the U.S.A., 1964, xiv, 334 p.

Thorndike, R.L.; Hagen, E. Measurement and evaluation in psychology and education. second ed. New York, John Wiley and Sons, Inc., 1961. 602 p., illus.

Tracey, W.R. Evaluation training and developemnt systems. New York, American Management Association, 1968. 304 p., bibliogr.

Tyler, R.W. Basic principles of curriculum and instruction: syllabus for Education 368. Chicago, The University of Chicago Press, 1950. 83 p.

_____. Changing concepts of educational evaluation. In Tyler, R.W. et al (eds.). Perspectives of curriculum evaluation, pp. 13-18. Chicago, Rand McNally, 1967. 102 p., illus. (AERA monograph series on curriculum evaluation, 1.).

_____. (ed.). Educational evaluation: new roles, new means. Chicago, The University of Chicago Press, 1969. xiii, 409 p. (68th Yearbook of the National Society for the Study of Education, part II.).

_____. The purposes of assessment. In: Beatty, W.H. (ed.). Improving educational assessment and an inventory of measures of affective behavior, pp. 2-13. Washington, D.C., Association for Supervision and Curriculum Development, N.E.A., 1969. viii, 164 p., bibliogr.

U.S. Civil Service Commission. An evaluation model designed to measure the effectiveness of a one-week training course "Position Classification and the Management Process". Washington, D.C., June 1970. 48 p. ED 064 568.

Van Gigch, J.P.; Hill, R.E. Using systems analysis to implement cost effectiveness and program budgeting in education. Englewood Cliffs, New Jersey, Educational Technology Publications, 1971. viii, 54 p., illus., bibliogr.

Veri, C.C. et al Evaluating teaching effectiveness in extension-administered programs. St. Louis, University of Missouri, Extension Division, 1973. 15 p. ED 078 275.

Warner, W.K. Problems in measuring the goal attainment of voluntary organizations. Journal of adult education (Phoenix, Arizona), vol. 19, no. 1. 1967.

Warr, P.; Bird, M.; Rackham, N. Evaluation of management training: a practical framework, with cases, for evaluating training needs and results. London, Gower Press, 1970. 111 p., illus. (A Gower Press special study.).

Wedemeyer, C.A. Evaluation of continuing education programs. American journal of pharmaceutical education (Lincoln, Nebraska), vol. 33, no. 5, May 1969, pp. 793-863.

Weiss, C.H. Evaluation research: methods for assessing program effectiveness. Englewood Cliffs, New Jersey, Prentice-Hall, Inc., 1972. xii, 160 p., bibliogr. (Prentice-Hall methods of social science series.).

Wilhelms, F.T. (ed.). Evaluation as feedback and guide. Washington, D.C., Association for Supervision and Curriculum Development, N.E.A., 1967. 282 p. (ASCD Yearbook, 1967).

Wittrock, M.; Wiley, D. (eds.). Evaluation of instruction: issues and problems. New York, Holt, Rinehart and Winston, 1970. xiii, 494 p., illus., bibliogr. (Symposium on problems in the evolution of instruction, University of California, Los Angeles, 1967.).

Woodward, J.P.; Yeager, J.L. Evaluation of programs to train educational R & D personnel. Pittsburgh, University of Pittsburgh, Learning Research and Development Center, April 1972. 26 p. ED 064 376.

Youmans, C.V. Testing for training and development. In: Craig, R.L.; Better, L.R. (eds.). Training and development handbook, pp. 55-86. New York, McGraw-Hill, 1967. xii, 650 p., illus., bibliogr.

APPENDIX 2

Examples of Questionnaires Used in
the Formative Evaluation of
a Course or Workshop

(A) Daily reaction form used in University of Illinois workshop on "Education and the Elderly". Reproduced by permission of the University of Illinois.

(B) Weekly reaction form used in UNISIST Course on Information Retrieval and Information Retrieval Systems, Katowice, Poland, August 1974.

2A

_____
Identification Number

DAILY REACTION FORM
Education and the Elderly

(The items on this form refer to workshop activities during the past 24 hours).

1. The following three items refer to the workshop sessions in which all participants met as a total group.

   a. Check the one phrase that best indicates how <u>important</u> and relevant the subject-matter and discussion topic was for you.

      5 ( ) Most important
      4 ( ) Very important
      3 ( ) Some importance
      2 ( ) Little importance
      1 ( ) No importance

   b. Check the one phrase that best indicates the extent to which the presentations had <u>clear</u> and to-the-point explanations.

      5 ( ) Exceptional
      4 ( ) Very good
      3 ( ) Good
      2 ( ) Fair
      1 ( ) Poor

   c. Check the one phrase that best indicates how understandable and useful the educational <u>materials</u> related to the total group sessions were.

      5 ( ) Exceptional
      4 ( ) Very good
      3 ( ) Good
      2 ( ) Fair
      1 ( ) Poor

2. Write in the following space the number of the small <u>sub-group</u> in which you are a participant.

   Group Number _____

3. Check the one phrase that best indicates how valuable your participation in this small sub-group has been for you.

5 ( ) Exceptional
4 ( ) Very good
3 ( ) Good
2 ( ) Fair
1 ( ) Poor

4. Check the one phrase that best indicates how satisfactorily you are progressing on your project, reading, and independent study.

5 ( ) Very well
4 ( ) Reasonably well
3 ( ) Somewhat
2 ( ) Hardly
1 ( ) Not at all

5. List below the most valuable new ideas related to the workshop that you have dealt with during the past 24 hours.

_____

_____

_____

_____

6. List below the main suggestions that you have which would have most improved your workshop experience during the past 24 hours.

_____

_____

_____

_____

7. Check the one phrase that best indicates your general rating of your workshop experience during the past 24 hours.

5 ( ) Really outstanding
4 ( ) Very satisfactory
3 ( ) Average
2 ( ) Just acceptable
1 ( ) Needs improvement

8. List below any other comments you have about the workshop at this point that might help to improve this and future workshops.

_____

_____

_____

_____

_____

_____

2B

Weekly Reaction Form

1. Check the one statement that best indicates how relevant the subject-matter of this week has been to you.

   Of great relevance ☐

   Of some relevance ☐

   Of little relevance ☐

   Of no relevance ☐

2. Check the one statement that best indicates your reaction to the quality of the lectures.

   Very good ☐

   Good ☐

   Average ☐

   Poor ☐

   Very Poor ☐

3. For each of the instructional methods mentioned below check the box that best describes your reaction to its value as a learning experience.

	Great value	Some value	Little value	No value
Transparencies				
MEDLARS slides				
Films				
MEDLINE demonstration				
Practical exercises				
Discussion periods				

4. Check the one statement that best describes how you feel you are progressing towards satisfying your own objectives in attending this course.

Great progress	☐
Good progress	☐
A little progress	☐
Hardly any progress	☐
No progress at all	☐

5. During this week what specific aspects of the course were <u>most</u> valuable in achieving your own objectives in attending?

6. During this week what specific aspects of the course were <u>least</u> valuable in achieving your own objectives in attending?

7. List any suggestions you have that you think might improve the course in later days or might improve the course if given again in the future.

8. Check the one statement that best indicates your general evaluation of the course during this week.

- Very good ☐
- Good ☐
- Average ☐
- Poor ☐
- Very poor ☐

9. If you have been able to complete some of the recommended supplementary readings for the course please indicate how valuable these readings have been to you so far.

- Very valuable ☐
- Valuable ☐
- Of little value ☐
- Of no value ☐

10. Select the statement that best indicates the novelty of the material presented this past week.

- Most of the material was new to me ☐
- About half the material was new to me ☐
- Less than half the material was new to me ☐
- None of the material was new to me ☐

11. Do you think the material presented in the past week could have been better presented some other way?

☐ No

☐ Yes What suggestions do you have for improved presentation?

APPENDIX 3

Examples of Questionnaires Used in the Summative Evaluation of a Course or Workshop

(A) Form used to evaluate the Eleventh Annual Clinic on Library Applications of Data Processing, University of Illinois, May 1974.

(B) End-of-course evaluation questionnaire used in the UNISIST Course on Information Retrieval and Information Retrieval Systems, Katowice, Poland, August 1974.

CLINIC QUESTIONNAIRE
Graduate School of Library Science
University of Illinois
Urbana, Illinois

Now that this Clinic is drawing to a close, we are certain that you have some reactions as to what aspects have been most valuable to you and what might have been different. As part of our effort to increase the effectiveness of future clinics, the planners would appreciate your completing this form. Your reactions will be confidential.

1. Please indicate your major purpose(s) for attending this Clinic.

   a. _____

   b. _____

   c. _____

   d. _____

2. What specific aspects of the Clinic experience contributed to your achieving one or any of the above personal objectives?

   _____

   _____

   _____

   _____

   _____

3. What specific aspects of the Clinic experience detracted from your achieving one or any of your previously stated objectives for attending the Clinic?

   _____

   _____

   _____

   _____

   _____

   _____

4. Would you give us your opinions in the following areas? (Check one in each area)

a. Content of presentations
- Excellent . . . . . 1( )
- Satisfactory. . . . . 2( )
- Unsatisfactory. . . . . 3( )

b. Number of presentations
- More needed . . . . . 1( )
- About right . . . . . 2( )
- Less needed . . . . . 3( )

c. Discussion and question period
- More time needed. . . . . 1( )
- About right . . . . . 2( )
- Less time needed. . . . . 3( )

d. Size of group at this Clinic
- Too many participants . . . . . 1( )
- Size about right. . . . . 2( )
- Too few participants. . . . . 3( )

e. Quality of presentations
- Excellent . . . . . 1( )
- Satisfactory. . . . . 2( )
- Unsatisfactory. . . . . 3( )

f. Level of treatment
- Above my head . . . . . 1( )
- About right . . . . . 2( )
- Too simple. . . . . 3( )

5. There are many parts of a Clinic experience that can either contribute to your satisfaction or detract from it. For each of the following, would you let us know how satisfied you have been. If any do not apply (e.g., you did not use hotel facilities), please leave blank.

a. Arrangements for coffee breaks
- Really outstanding. . . . . 1( )
- Very satisfactory . . . . . 2( )
- Average . . . . . 3( )
- Just acceptable . . . . . 4( )
- Need improvement. . . . . 5( )

b. Hotel Rooms
- Really outstanding. . . . . 1( )
- Very satisfactory . . . . . 2( )
- Average . . . . . 3( )
- Just acceptable . . . . . 4( )
- Need improvement. . . . . 5( )

c. Meeting Rooms

Really outstanding. . . . . 1( )
Very satisfactory . . . . . 2( )
Average . . . . . 3( )
Just acceptable . . . . . 4( )
Need improvement. . . . . 5( )

d. Organized social events (i.e., opening dinner and social hour)

Really outstanding. . . . . 1( )
Very satisfactory . . . . . 2( )
Average . . . . . 3( )
Just acceptable . . . . . 4( )
Need improvement. . . . . 5( )

6. In your opinion, how well did the total set of presentations cover the topic designated for this particular Clinic? (Check one)

Comprehensive Coverage					Inadequate Coverage
5	4	3	2	1	

7. In comparison with other continuing education activities which you have attended during the past few years, how would you rate this Clinic?

Excellent					Very Poor
5	4	3	2	1	

8. Prior to your attending, in what ways did you hear about this Clinic? (Check all that apply)

Received a brochure through the mail. . . . . 1( )
Saw news item in journals . . . . . 2( )
Through a friend or acquaintance. . . . . 3( )
Other_____ . . . . . 4( )
(specify)

9. Please read all of the following statements. Then, check any of those that state how you feel about the Clinic as a whole.

It was one of the most rewarding experiences I have ever had . . . . . 1( )
Exactly what I wanted . . . . . 2( )
It provided the kind of experience I can apply to my own situations. . . . . 3( )
It helped me personally . . . . . 4( )
It solved some problems for me. . . . . 5( )
I think it served its purpose . . . . . 6( )
It has some merits. . . . . 7( )
It was fair . . . . . 8( )
It was neither very good nor very poor. . . . . 9( )
I was mildly disappointed . . . . . 10( )

(continued on next page)

It was not exactly what I needed. . . . . 11( )
It was too general. . . . . 12( )
I am not taking any new ideas away. . . . . 13( )
It didn't hold my interest. . . . . 14( )
It was too superficial. . . . . 15( )
It was very poorly planned. . . . . 16( )
I didn't learn a thing. . . . . 17( )
It was a complete waste of time . . . . . 18( )

10. Would you be interested in attending another Clinic of this kind next year?

_____Yes

_____No

11. What topics do you recommend for next year's Clinic?

a. _____

b. _____

c. _____

d. _____

Thank you for your cooperation!

- 3B -

Unesco

International Training Course
Katowice, Poland
August 1974

END-OF-COURSE EVALUATION

1. When you decided to attend this course you had certain objectives regarding what you hoped to gain from it. Circle the statement that best describes how far the course has met these objectives.

5	4	3	2	1
Completely	Almost completely	Partially	Incompletely	Very incompletely

2. In terms of your own professional goals indicate how valuable this course is likely to be to you when you return to your own country.

5	4	3	2	1
Extremely valuable	Quite valuable	Of some value	Of little value	Of no value at all

3. Circle the one statement that best describes your assessment of the quality of the instruction as a whole.

5	4	3	2	1
Very good	Good	Average	Poor	Very poor

4. Would you please give us your evaluation on each of the following aspects? Check one box for each aspect.

(a) Discussion and question periods

More time needed ☐

About right ☐

Less time needed ☐

(b) Student practical sessions

More time needed ☐

About right ☐

Less time needed ☐

(c) Size of course

Too many participants ☐

Size about right ☐

Too few participants ☐

(d) Level of treatment

Subject-matter too complex ☐

Treatment about right ☐

Treatment too simple ☐

5. There are many factors that may contribute to the satisfaction of participants in an educational experience of this type. For each of the following factors would you indicate how satisfied you have been?

(a) Arrangements for morning and afternoon breaks

Very good ☐

Satisfactory ☐

Unsatisfactory ☐

(b) Hotel accommodations

Very good ☐

Satisfactory ☐

Unsatisfactory ☐

(c) Lecture rooms

Very good ☐

Satisfactory ☐

Unsatisfactory ☐

(d) Organized social events

Very good ☐

Satisfactory ☐

Unsatisfactory ☐

6. In your opinion how well did the complete sequence of lectures cover the subject of information retrieval?

Comprehensive coverage ☐

Satisfactory coverage ☐

Inadequate coverage ☐

7. Select the statement that best indicates the novelty of the material presented in this course.

Most of the material was new to me ☐

About half was new to me ☐

Less than half was new to me ☐

None of the material was new to me ☐

8. If another Unesco course of this type is conducted in the future would you recommend it to your colleagues?

YES ☐

NO ☐

9. If Unesco organizes a more advanced course in this subject area would you be interested in attending?

YES ☐

NO ☐

10. List below any suggestions you have as to how this course might have been improved.

11. Do you know what is meant by the term controlled vocabulary as applied to an information retrieval system?

☐ NO

☐ YES What are the major functions of a controlled vocabulary in information retrieval?

12. Do you understand the difference between an off-line computer system and an on-line computer system?

☐ NO

☐ YES What are the major advantages of an on-line system over an off-line system for purposes of information retrieval?

13. With what developments or activities in information retrieval do you associate the following names:

Name	Development
William Batten	_____
Calvin Mooers	_____
Mortimer Taube	_____
Eugene Garfield	_____
Hans Peter Luhn	_____

14. What is meant by selective dissemination of information (SDI)?

15. How does SDI differ from retrospective searching?

16. What is a machine-readable data base?

17. What is the Science Information Exchange?

18. What is a thesaurus as applied to information retrieval?

19. How are hierarchical relationships displayed in a thesaurus of this type?

20. By what criteria do you think that users will judge the effectiveness of an information retrieval service?

21. What is a citation index?

22. What is an "invisible college"?

APPENDIX 4

Pre-course Questionnaire Used in
Evaluation of Learning
Occurring in the UNISIST Course on
Information Retrieval and Information
Retrieval Systems, Katowice, Poland,
August 1974

PRE-COURSE QUESTIONNIARE

1. Describe briefly the reasons that made you decide to enrol in this course.

2. Describe briefly your major expectations regarding what you hope to gain from this course.

3. Do you know what is meant by the term controlled vocabulary as applied to an information retrieval system?

☐ NO

☐ YES What are the major functions of a controlled vocabulary in information retrieval?

4. Do you understand the difference between an off-line computer system and an on-line computer system?

   ____ NO

   ____ YES What are the major advantages of an on-line system over an off-line system for purposes of information retrieval?

5. With what developments or activities in information retrieval do you associate the following names:

Name	Development
William Batten	_____
Calvin Mooers	_____
Mortimer Taube	_____
Eugene Garfield	_____
Hans Peter Luhn	_____

6. What is meant by selective dissemination of information (SDI)?

7. How does SDI differ from retrospective searching?

8. What is a machine-readable data base?

9. What is the Science Information Exchange?

10. What is a thesaurus as applied to information retrieval?

11. How are hierarchical relationships displayed in a thesaurus of this type?

12. By what criteria do you think that users will judge the effectiveness of an information retrieval service?

13. What is a citation index?

14. What is an "invisible college"?

APPENDIX 5

Report on the Evaluation of the Eleventh Annual Clinic on Library Applications of Data Processing, University of Illinois, May 1974

F.W. Lancaster

July 1974

University of Illinois

Graduate School of Library Science

Introduction

The eleventh annual Clinic on Library Applications of Data Processing was held by the Graduate School of Library Science, University of Illinois, from April 28 to May 1, 1974. The subject of this clinic, which was attended by approximately 140 librarians and others concerned with data processing applications, was minicomputers and their use in libraries. An attempt was made to evaluate the success of the clinic by the use of a questionnaire administered to all participants attending the final session of the clinic in May. The questionnaire used is reproduced in the appendix to this report. Questionnaires were received from 79 participants and the responses recorded are summarized in the remainder of this report.

Reasons for attending the clinic

The majority of the respondents (42/79) indicated that they attended the clinic in order to obtain a state-of-the-art review on the use of minicomputers in libraries. However, a substantial number (23/79) said that their main reason for attending was to meet other people working in the field, to find out what these people are doing, and to discover their opinions on various aspects of the topic of the clinic.

Major factors contributing to the attainment of objectives

The aspect of the clinic most frequently mentioned as contributing to the attainment of the personal objectives of participants was "conversations with those attending", mentioned by 25 of the respondents. As far as the formal aspects of the clinic are concerned, the group of papers from the University of Minnesota was the session most frequently mentioned (by 21 of the respondents). Mentioned six times each were the Curley presentation, the Beaumont presentation, the Davison presentation, and the tutorial by Corey and Divilbiss. "Social events" was mentioned five times by respondents and this could be grouped with the category "conversations with those attending" since, presumably, the main value of the social events lies in the opportunity provided for conversation with other participants.

Factors detracting from the attainment of objectives

The factor most frequently mentioned as a detraction was the quality and number of the audiovisual aids: fifteen respondents mentioned that these were inadequate or of poor quality. The remaining criticisms were conflicting. For example, a number of people said that there were too many "how to do it" talks and an approximately equal number said there were too few papers of this type and too many "promotional" presentations (this despite the fact that only two of the papers were given by representatives of commercial organizations).

Nine people suggested that the clinic would be better if the papers were available in advance. Five people suggested that the evening presentations should be eliminated to allow more time for informal contacts among participants.

Content of presentations

Excellent	21	28%
Satisfactory	42	56%
Unsatisfactory	0	
Variable (some satisfactory, some not)	12	16%
Nonrespondents	4	
TOTAL	79	100

Number of presentations

More needed	7	10%
About right	61	84%
Less needed	5	7%
Nonrespondents	6	
TOTAL	79	101

Discussion and question period

More time needed	16	22%
About right	57	77%
Less time needed	1	1%
Nonrespondents	5	
TOTAL	79	100

Quality of presentations

Excellent	15	20%
Satisfactory	44	59%
Unsatisfactory	4	5%
Quality varied	12	16%
Nonrespondents	4	
TOTAL	79	100

Level of treatment

Above my head	7	9%
About right	50	67%
Too simple	4	5%
Level varied	14	19%
Nonrespondents	4	
TOTAL	79	100

Coffee breaks

Really outstanding	7	9%
Very satisfactory	37	49%
Average	26	35%
Just acceptable	2	3%
Need improvement	3	4%
Nonrespondents	4	
TOTAL	79	100

Hotel rooms

Really outstanding	20	28%
Very satisfactory	36	50%
Average	14	19%
Just acceptable	2	3%
Need improvement	0	
Nonrespondents	7	
TOTAL	79	100

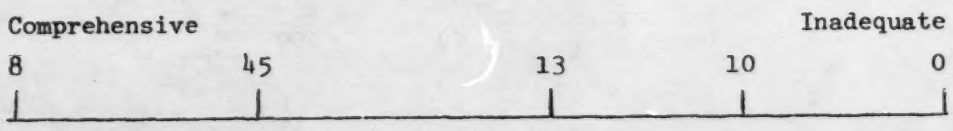
Meeting rooms

Really outstanding	10	13%
Very satisfactory	33	43%
Average	26	34%
Just acceptable	5	6%
Need improvement	3	4%
Nonrespondents	2	
TOTAL	79	100

Organized social events

Really outstanding	5	7%
Very satisfactory	33	46%
Average	23	32%
Just acceptable	9	12%
Need improvement	2	3%
Nonrespondents	7	
TOTAL	79	100

Extent of coverage of clinic topic



Comparison with other continuing education activities attended in the past few years

Excellent				Very poor
17	37	14	3	0

How the attendee heard about the clinic

Received a brochure through the mail	41
Saw news item in journal	36
Through a friend or acquaintance	20
Other	15
Nonrespondents	2
	114*

*These are not mutually exclusive categories

General reaction to the clinic (Kropp-Verner Scale)

One of the most rewarding experiences I have ever had	4	5%
Exactly what I wanted	13	17%
Provided the kind of experience I can apply to my own situation	39	51%
It helped me personally	35	45%
It solved some problems for me	24	32%
I think it served its purpose	42	54%
It has some merits	12	16%
It was fair	2	3%
It was neither very good nor very poor	7	9%
I was mildly disappointed	14	18%
It was exactly what I needed	7	9%
It was too general	2	3%
I am not taking any new ideas away	1	1%
It didn't hold my interest	1	1%
It was too superficial	1	1%
It was very poorly planned	0	
I didn't learn a thing	0	
It was a complete waste of time	0	
Nonrespondents	2	
	206*	**

*The categories are not mutually exclusive

**These are percentages of the 77 respondents who checked each answer.

Would participant attend another clinic next year?

Yes	70	95%
No	4	5%
Nonrespondents	5	
TOTAL	79	100

Suggested topics for future clinics

Management, staffing, training	10
Cooperation and networking	10
Costs and economics	8
Systems for smaller libraries	7
Description and evaluation of operational systems	7
Patron use, reaction and training	6
Information retrieval	6
Serials control	5
New hardware	4
Data bases and data base services	4
Processing centers and services	3
Standards for automation	3
Changing role of the librarian	2

Conclusions and recommendations

On the whole it seems fair to say that this clinic was successful and that the great majority of those attending were well satisfied with the content and organization. There are many indications, however, that the informal aspects of a meeting of this type are at least as important as the formal aspects. A fairly substantial group of respondents indicated the importance of the social functions and the value of meeting with others working in the field. Others indicated a need to provide more time for discussions on an informal basis. In view of this reaction I suggest that we eliminate formal evening sessions at the next clinic except for the opening session on Sunday evening. The other two evenings will either be completely free or will be reserved for demonstrations or discussion periods among smaller groups of participants.

It also seems clear that we should try to exert more influence over participants in terms of their use of audiovisual materials. Although some were very good in this clinic (e.g., most of the BIBNET material), others were poor. In future I propose, when inviting further speakers, to stress the importance of good audiovisuals and to point out the size of type needed for adequate legibility in a group of the size we can expect.

A number of attendees expressed a desire for the papers to be printed and available in advance of the clinic or, at least, at the time the clinic is held. I am somewhat lukewarm about this idea. It is hard enough to get people to agree to present a paper and then to meet our publication deadline. It would, I feel, be much more of a problem to get them to prepare a formal paper in advance of the meeting. Moreover, it has been my experience that, in a large meeting, only a small proportion of those attending have read the papers in advance. Pre-distribution of papers works well with a small conference of invited participants

(about 25 to 30 people) but is a less attractive proposition, I feel, for a group of over a hundred.

There is little evidence, from the results of this survey, that the size of the audience was considered too large. I would therefore suggest that we have no real need to restrict registration to a hundred, but should be able to go up to 125 or 150 without adverse effect on the clinic.

Finally, I believe that this formal evaluation has been useful and that we should continue to evaluate future clinics using approximately the same questionnaire. This will give us some standard of comparison from year to year.

APPENDIX 6

FOLLOW-UP QUESTIONNAIRE
ON UNISIST INTERNATIONAL TRAINING
COURSE ON INFORMATION RETRIEVAL
AND INFORMATION RETRIEVAL SYSTEMS
KATOWICE, POLAND
AUGUST 1974

The course you attended in Katowice some months ago can be considered to have had a number of separate but closely related objectives:

1. To present a broad overview of the field of information retrieval with special reference to procedures used in modern information services.

2. To give participants a better understanding of what subject-matter needs to be covered in a curriculum in the field of information retrieval.

3. To illustrate in detail how a particular course in this subject might be conducted, including considerations of subject content, organization, teaching methods and teaching materials.

4. To stimulate participants to give more serious consideration to the problems of information retrieval in developing countries and to ways in which modern information handling systems might be introduced into these countries.

We would like your retrospective view of how well the Katowice course fulfilled these various objectives. Circle a number on each scale to indicate your assessment of the degree to which the course satisfied each of the four objectives (the higher numbers indicate greater satisfaction).

	Course completely satisfied this objective ↑						Course contributed nothing to this objective ↓
Objective 1:	7	6	5	4	3	2	1
Objective 2:	7	6	5	4	3	2	1
Objective 3:	7	6	5	4	3	2	1
Objective 4:	7	6	5	4	3	2	1

We would now like you to consider in what ways this particular UNISIST course has contributed to you professionally. Please indicate in what way, if any, you have been able to make use of the learning acquired during this course. Please check all appropriate boxes.

Have drawn upon the course in the planning of educational programs in my own country ☐

Have drawn upon the course in the development of a curriculum in the information sciences area ☐

Have used material gained from the course in my own teaching ☐

Was able to draw upon information gained from the course in the development of information services in my own country ☐

Have benefited from the course in other ways ☐
Please explain:

Have been unable to make any use of the material presented in this course ☐
Please explain why:

Please indicate on the scale below how valuable you feel this course has been to you professionally:

Great value						No value
7	6	5	4	3	2	1

If you have any further comments to make on the Katowice course please record these below.

Unesco is very grateful for your co-operation in completing this brief questionnaire. Please return it when completed to:

APPENDIX 7

Example of a Checklist for Use
by an Independent Observer

Reproduced from:

U.S. Civil Service Commission. Bureau of Training. Training Evaluation: a Guide to its Planning, Development, and Use in Agency Training Courses. Washington, D.C., May 1971. 21 p. Pamphlet T-13. ED 054 365.

CLASSROOM OBSERVATION GUIDE

Course_____ Session_____

Time_____ Date_____ Instructor_____

1. Training objectives: _____

2. Assessment: To what extent did training objectives appear to be attained? (Adverse factors should be explained in item 4 below.)

   a. ☐ Fully. (Minor deficiencies, if any.)

   b. ☐ Partially. (Some factors preventing full achievement.)

   c. ☐ Hardly at all.

3. Checks on attainment: What steps did the instructor take to assure himself that training objectives were being attained? _____

   How effective were they?_____

4. Factors influencing objective attainment: Factors that had an adverse effect on training objective attainment as indicated in 2 above. (Check any such factors and explain on the back of this sheet.)

Factor		Check
A.	Introduction to training session	☐
B.	Content	☐
C.	Clarity of explanations	☐
D.	Instructor's apparent knowledge of subject	☐
E.	Subject-matter organization	☐
F.	Selection and use of training aids	☐
G.	Instructor's manner (tact, etc.)	☐
H.	Instructor's ability to maintain trainees' interest	☐
I.	Instructor's ability to direct discussion	☐
J.	Appropriateness of material to group	☐
K.	Time/material relationship	☐
L.	Individual trainee involvement	☐

Factor (cont.)		Check
M.	Instructional methods used	____
N.	Reference materials	____
O.	Classroom facilities	____
P.	Tests and practical exercises	____
Q.	Other (specify)	____

_____
(Observer's Name)

Index of First Occurrence of Special Terms

	Page
Antecedents	11
Closed-ended question	44
Cost-benefit analysis	4
Cost-benefit evaluation	4
Cost effectiveness evaluation	4
Course	3
Effectiveness evaluation	4
"Evaluation overkill"	35
Formative evaluation	8
Kropp-Verner scale	53
Macro-evaluation	4
Micro-evaluation	4
Multiple choice questions	45
Outcomes	11
Open-ended question	44
Programme objectives	33
Ranking scale	53
Semantic differential scale	53
Seminar	3
Summative evaluation	8
Thurstone scale	47
Transactions	11
Workshop	3